NOTE TO USERS

The original manuscript received by UMI contains pages with slanted print. Pages were microfilmed as received.

This reproduction is the best copy available

UMI

University of Alberta

Genetic Mapping of DNA Sequences Homologous to Plant Disease Resistance Genes in *Brassica napus* L.

by

Alexandre Joyeux ©

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for
the degree of Master of Science

In

Molecular Biology and Genetics

Department of Biological Sciences

Edmonton, Alberta

Spring 1998

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-28953-2

Canada

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "Genetic Mapping of DNA Sequences Homologous to Plant Disease Resistance Genes in *Brassica napus* L." submitted by Alexandre Joyeux in partial fulfillment of the requirements for the degree of Master of Science in Molecular Biology and Genetics.

Dr. Allen G. Good

Dr. Ross Hodgetts

Dr. Alan Grombacher

Dr. J.P. Tewari
(External Examiner)

Abstract

We present the genetic mapping of *Brassica napus* DNA sequences amplified with primers corresponding to both the nucleotide-binding site and the leucine-rich repeat domain of the *Arabidopsis thaliana RPS2* gene. To our knowledge, the sequence conservation in the leucine-rich repeat domain of known resistance genes has not yet been exploited via PCR. Two of our sequences are highly homologous to plant disease resistance genes, and other sequences are similar to animal genes involved in other stages of the disease response. The sequences map both singly and in clusters throughout the *B. napus* genetic map, which is characteristic of known resistance loci. We have also designed a minimal *Brassica napus* mapping population of twenty-three doubled-haploid individuals. We have validated this method and used this population in mapping our resistance gene homologues. In addition, we introduce a method for the quick mapping of resistance gene homologues using the polymerase chain reaction.

Acknowledgements

I would like to thank my supervisor, Dr. Allen Good, for his guidance and support throughout this degree. I would also like to thank the other members of my committee, Dr. Ross Hodgetts and Dr. Alan Grombacher for their helpful advice and support, as well as Dr. J.P. Tewari for his time as my external examiner.

Thank you to Dr. Marc Fortin, from McGill University, for his moral and intellectual support during my undergraduate degree, and throughout my M.Sc. I am also grateful for his technical advice in the development of the method for quick mapping of disease resistance gene homologues.

Thank you also to my lab colleagues and friends, Reinhold, Susan, Christie, James, Bret, Mary, Usukuma, and Nick for being there, and for helping me enjoy my time in and outside the lab.

We are very grateful to Dr. Tom Osborn, from University of Wisconsin, for generously providing the mapping population, as well as the mapping dataset.

Special thanks to Dr. Shauna Sommerville for fascinating discussions about plant disease resistance, when she visited the department in October 1997.

Finally, as a graduate degree is much more than research, I would like to thank my very close friends Trista Wagoner, Amy Dunham, Nadia Bhuiyan, Rebecca Tittler, Susan Kiesling, and Catherine Beaudoin for being there, through the good times and the bad, for making me laugh once in a while, for listening to me talk all the time, and for giving me the strength that we all need to go through life and make our dreams come true.

Financial support was generously provided by the Natural Sciences and Engineering Research Council (NSERC), and by the University of Alberta.

Table of Contents

1.0 General Introduction ____ *1*

1.1 Plant Pathology Through the Ages ____ **1**
1.1.1 Famous Epidemics ____ 1
1.1.2 Controlling Plant Pathogens: from Gods to Genes ____ 3
1.1.3 The Gene-for-Gene Theory ____ 5
1.1.3.1 Pathogen Avirulence Genes ____ 6
1.1.3.2 Plant Disease Resistance Genes ____ 8
1.1.3.3 Modern Evidence ____ 13
1.1.3.4 Exceptions ____ 16
1.1.3.5 Homology to Animal Systems ____ 17

1.2 Isolating Plant Disease Resistance Genes ____ **21**
1.2.1 Transposon tagging ____ 22
1.2.2 Genetic Mapping and Map-Based Cloning ____ 24
1.2.2.1 High Resolution Plant Genetic Mapping, Pooled Sampling, and Bulked Segregant Analysis ____ 28
1.2.2.2 Comparative Genome Mapping ____ 32
1.2.3 Mapping of Plant Disease Resistance Gene Homologues ____ 33
1.2.4 Random Genome Sequencing ____ 36

1.3 Summary ____ **37**

1.4 Objectives ____ **37**

1.5 Literature Cited ____ **39**

2.0 Genetic Mapping of Plant Resistance Gene Homologues in Brassica napus L., using a minimal mapping population ____ *49*

2.1 Introduction ____ **49**

2.2 Materials and Methods ____ **52**
2.2.1 Plant materials ____ 52
2.2.2 Genomic DNA Extractions ____ 52
2.2.3 Amplification and subcloning of disease resistance gene homologues ____ 53
2.2.4 DNA Sequencing ____ 55
2.2.5 Southern hybridizations ____ 55
2.2.6 Quick genetic mapping using the polymerase chain reaction ____ 56
2.2.7 Linkage analysis ____ 56
2.2.8 Selection and validation of the minimal mapping population ____ 57

2.3 Results ____ **58**
2.3.1 Resistance gene homologues in the Brassicaceae ____ 58
2.3.2 Subcloning of *Brassica napus* Resistance Gene Homologues ____ 61
2.3.3 Validity of minimal mapping population ____ 64
2.3.4 Genetic mapping ____ 65
2.3.5 Sequence Analysis ____ 70
2.3.6 Quick mapping using the polymerase chain reaction ____ 80

2.4 Discussion ____ **85**

2.5 Literature Cited ____ **94**

3.0 General Discussion ____ *98*

3.1 Isolation of Disease Resistance Gene Homologues in Plants ____ **98**
3.2 Use of Minimal Populations in Genetic Mapping ____ **100**
3.3 Quick Mapping of Putative Disease Resistance Genes ____ **101**
3.4 Future Prospects ____ **101**
3.5 Concluding Remarks ____ **103**
3.6 Literature Cited ____ **103**
Appendix 1: Buffers and Solutions ____ ***105***

List of Tables

Table 1.1 Cloned plant disease resistance genes _____ 9

Table 2.1 Primers for amplification of resistance gene homologues in the Brassicaceae _____ 54
Table 2.2 PCR reaction conditions for the amplification of R gene homologues ____ 54
Table 2.3 Number of clones and clone families obtained _____ 61
Table 2.4 Three main classes of clones identified by their *Rsa*I pattern _____ 62
Table 2.5 Grouping of polymorphic clones into eight classes _____ 66
Table 2.6 Genotypes of the minimal mapping population for the six polymorphic clones _____ 69
Table 2.7 Homology search results for the ten sequences studied _____ 78
Table 2.8 Genotypes of the mapping population for PCR-generated markers _____ 83

List of Figures

Figure 1.1 Illustration of the gene-for-gene concept ____ 6
Figure 1.2 Homologies between plant and animal disease resistance pathways ____ 18

Figure 2.1 Amplification of R gene homologues in Arabidopsis and *Brassica napus* ____ 59
Figure 2.2 Amplification of R gene homologues in members of the Brassicaceae ____ 60
Figure 2.3 Number of clones per family, for primers PLRR-1 ____ 62
Figure 2.4 Number of clones per family, for primers PLRR-2 ____ 62
Figure 2.5 *Rsa*I restriction patterns obtained for twenty-seven clone families ____ 63
Figure 2.6 Genetic distances and LOD scores for first five markers on the *Brassica napus* genetic map ____ 65
Figure 2.7 Genetic distances and LOD scores for sixth marker on the *Brassica napus* genetic map ____ 65
Figure 2.8 Agarose gel of the ten polymorphic clones used in our mapping studies ____ 67
Figure 2.9 Southern blot showing polymorphic marker P1-15B ____ 68
Figure 2.10 Sequence alignment of inserts P1-2A, P1-3A, and P1-13A ____ 72
Figure 2.11 Partial sequence of insert P1-16A ____ 73
Figure 2.12 Complete sequence of insert P2-18A ____ 73
Figure 2.13 Complete sequence of insert P1-7B ____ 74
Figure 2.14 Partial sequence of insert P1-15B ____ 74
Figure 2.15 Partial sequence of insert P2-13B ____ 75
Figure 2.16 Complete sequence of insert P2-20B ____ 75
Figure 2.17 Complete sequence of insert P1-11C ____ 76
Figure 2.18 Amino acid sequence alignment of conserved R gene domains, for *RPS2* and the clones P1-16A and P2-13B ____ 79
Figure 2.19 Agarose gel showing the testing of parental lines in the quick mapping of R gene homologues using PCR ____ 81
Figure 2.20 Agarose gel showing the testing of the minimal mapping population in the quick mapping of R gene homologues using PCR ____ 82
Figure 2.21 *Brassica napus* genetic map incorporating the 16 markers mapped in this study ____ 84
Figure 2.22 Three possible map positions of marker P2-13B-Hb ____ 92

Abbreviations

AFLP	-	**A**mplified **F**ragment **L**ength **P**olymorphism
Avr	-	**Avi**rulence
bp	-	**B**ase **P**air
cDNA	-	**C**omplementary **DNA**
cfu	-	**C**olony **F**orming **U**nit
cM	-	**C**enti**M**organ
CMS	-	**C**ytoplasmic **M**ale **S**terility
CTAB	-	**H**exade**C**yl**T**rimethyl**A**mmonium **B**romide
dCTP	-	**D**eoxy**C**ytosine **T**riphosphate
DNA	-	**D**eoxyribo**N**ucleic **A**cid
DH	-	**D**oubled **H**aploid
EDTA	-	**E**thylene**D**iamine**T**etraacetic **A**cid
EST	-	**E**xpressed **S**equence **T**ag
g	-	**G**ram
HR	-	**H**ypersensitive **R**esponse
Hrp	-	**H**ypersensitive **R**esponse and **P**athogenesis
iPCR	-	**I**nverse **PCR**
IPTG	-	**I**so**P**ropyl-β-D-**T**hio**G**alactopyranoside
kb	-	**K**ilo **B**ase **P**air
LG	-	**L**inkage **G**roup
LRR	-	**L**eucine-**R**ich **R**epeat
MAS	-	**M**arker-**A**ssisted **S**election
NBS	-	**N**ucleotide-**B**inding **S**ite
NIL	-	**N**ear-**I**sogenic **L**ine
PCR	-	**P**olymerase **C**hain **R**eaction
P-loop	-	**P**hosphate-binding **Loop**
PLRR	-	**P**-loop/**L**eucine-**R**ich **R**epeat
PR	-	**P**athogenesis-**R**elated
RAPD	-	**R**andom **A**mplified **P**olymorphic **D**NA
RFLP	-	**R**estriction **F**ragment **L**ength **P**olymorphism
RGA	-	**R**esistance **G**ene **A**nalogue
R-gene	-	Disease **R**esistance **Gene**
SA	-	**S**alycilic **A**cid
SAR	-	**S**ystemic **A**cquired **R**esistance
SDS	-	**S**odium **D**odecyl **S**ulfate
SSC	-	**S**alt **S**odium **C**itrate
T_m	-	**T**emperature
U	-	Unit(s)
(v/v)	-	**V**olume per **V**olume
(w/v)	-	**W**eight per **V**olume
Xgal	-	5-bromo-4-chloro-3-indolyl-β-D- **gal**actopyranoside

1.0 General Introduction

1.1 Plant Pathology Through the Ages

1.1.1 Famous Epidemics

Humans have depended on plants in virtually every aspect of their lives since the dawn of times. Whether it is for the oxygen we breath, for the food we eat, for the clothes we wear, for the medication we take, for the coffee we drink, plants play a key role in our lives.

Ever since mankind switched from a nomadic to a more sedentary mode of life, approximately 10000 years ago, we have learned to use to our advantage plants' high capacity for adaptation by selecting the plants which were the most adapted to our environment, and which suited our needs best. This resulted in high yielding, high performance, and often very homogeneous plant varieties that were very different from their wild ancestors. The early development and success of entire civilizations has largely depended on mankind's successes in plant domestication. It is therefore not surprising that large plant disease epidemics, like human disease epidemics, have had disastrous consequences for entire societies, and have often shaped the course of history.

The first known written account of plant diseases dates back to the Greek philosopher Theophrastus (370-286 B.C.) who realized that diseases were more severe in lowlands, where it was more humid than on the hillsides, and that some diseases preferred particular crops (Agrios, 1988). Until the invention of the microscope in the 17th century, however, the causal agents of plant diseases were not known, and were,

more often than not, attributed to evil forces rather than to other living organisms (Agrios, 1988).

Perhaps the most famous plant disease epidemic to shape history was the great Irish potato famine of 1846. Potato was a staple crop in Northern Europe, and many families depended on it for their survival. Between the exceptionally humid summer of 1845, which favored the germination and proliferation of the disease-causing fungus, and the fact that potatoes were grown everywhere, and in large uniform stands, the disaster was unavoidable. The epidemic that struck Ireland devastated virtually the entire potato crop. The outcome: one million deaths, and more than one and a half million people were forced to leave Ireland and to emigrate to the United-States (Agrios, 1988).

Though this episode was particularly tragic, and though it stimulated a large interest in the study of plant diseases, it was by far not the last. Other examples include the downy mildew of grapes which, after being introduced from the United-States in 1878, threatened Europe's wine industry (Agrios, 1988). The coffee rust disease also had its impact in Ceylon, where it devastated the entire crop at the end of the 19th century (Agrios, 1988) and forced the British colonists to implement tea as an alternative cash crop.

Despite all the modern agricultural technologies and our increasing knowledge of plant pathology, plant disease epidemics still cause extensive damage throughout the world, year after year. A recent and infamous example is the 1970 epidemics of southern corn leaf blight and yellow corn leaf blight in the United States of America. One of the most successful modes of corn breeding is the use of two very inbred lines with complementary favorable characteristics to produce a hybrid which is of higher quality

than either parent. This requires that one of the parents be male sterile. In 1970, over 85% of all US corn hybrids were produced using the T (Texas) source of cytoplasmic male sterility (cms) (Pring and Lonsdale, 1989). It turned out that this source of cms also caused acute susceptibility to the toxin produced by certain races of corn blight fungi. Approximately 17% of the US corn crop, and one billion dollars, were lost in one summer (Pring and Lonsdale, 1989).

The causes for all these epidemics were multiple, but one common factor was the monoculture of similar varieties of a same crop over large surface areas. This not only leaves the crop particularly vulnerable to disease development and propagation, but it also strongly selects for pathogens that have the ability to attack the specific variety grown. Scientists were already warning against monoculture during the coffee rust epidemic, when they observed that the disease was more destructive in Ceylon's uniform coffee stands than in mixed stands, such as those grown in Brazil (Agrios, 1988). The expanding world population, increasingly specialized agricultural techniques, and rapidly escalating economic pressures, however, favor large scale production of a few selected staple crops. Hence, the world depends largely on rice, wheat, maize, and potatoes for its food supply. These and many other cash crops are grown over large uniform areas, leaving them highly vulnerable to pest epidemics.

1.1.2 Controlling Plant Pathogens: from Gods to Genes

The Roman Empire had recognized the destructive power of plant diseases, cereal rusts in particular, and had created a god, Robigo, to protect their crops. "Crop protection" then referred to the sacrifice of red sheep and dogs for Robigo, every spring,

to keep the rust diseases away and to spare the crops (Agrios, 1988). Until the 19th century, when the theory of spontaneous generation was put to rest, plant diseases were not well understood, and crop protection remained largely dependent on invocations and sacrifices (Agrios, 1988).

It was only in 1853 that DeBary established once and for all that the fungi he was working with were the cause rather than the result of plant diseases (Agrios, 1988). Despite the description and more general understanding of plant disease causing organisms, it wasn't until the 1900s that effective chemical treatments were developed. Pesticides came as a blessing in many areas of the world and were used extensively throughout the 1st half of the century. Their use is still widespread today, though the high toxicity of many of these chemicals is more widely recognized and they are subjected to a certain level of regulation. It became apparent in the 1970s however that many pathogens were evolving resistance to protective and systemic fungicides (Agrios, 1988). This time also marks the beginning of an era of increased awareness as to the need for more sustainable agricultural practices. This awareness was initiated by a general reflection on the Green Revolution of the 1960s which, though it brought undeniable yield increases in major crops, was hardly sustainable.

As for plant disease resistance, though it was recognized, as mentioned before, that certain pathogens only affected certain crops, it was not until 1905 (as Mendel's findings were rediscovered) that Biffen showed that resistance to a rust fungus was inherited in a Mendelian fashion in two wheat varieties (Biffen, 1905). This was followed, in 1914, by Stakman's discovery that pathogens have physiological races that differ in their ability to infect different plant genotypes (Stakman, 1914). This discovery

eventually led to the gene-for-gene concept, which establishes that for every disease resistance gene in the plant, there is a corresponding avirulence gene in the pathogen (Flor, 1946). Finally, Vanderplank (1963) distinguished between vertical resistance, conferred by one or few major gene(s), and horizontal disease resistance, conferred by many minor genes. The discovery that plants possess genes that allow them to fight against their pathogens, opened a whole new range of possibilities for crop protection.

1.1.3 The Gene-for-Gene Theory

Flor (1946), in an extensive study of the genetics of flax-rust interactions, discovered that resistance or disease were the result of a specific interaction between "factors" in the rust fungus and corresponding specific "factors" in the host. It was also found that combining different factors by hybridizing two fungus races did not increase virulence on flax if the host was resistant to both parental races. Finally, Flor observed that the F_2 hybrids of a virulent and an avirulent race of the fungus segregated in a monofactorial ratio if the host variety possessed one "pair of factors" conferring resistance to the avirulent parent, in a bifactorial ratio if it possessed two pairs of resistance factors, and in a trifactorial ratio if it possessed three pairs (Flor, 1946; and reviewed in Flor, 1971). This led Flor to hypothesize that, for every pair of resistance factors in the host, there is a corresponding avirulence pair in the fungus. This became known as the gene-for-gene concept, which states that the presence of a gene in one population depends on the presence of a corresponding gene in another population, and that the interaction of these two genes produces a single phenotype (Person *et al.*, 1962). In practice, this means that disease resistance occurs only upon the interaction of the

products of a dominant resistance gene allele in the plant, with the corresponding dominant avirulence gene in the pathogen (Figure 1.1) (Flor, 1971).

	Genes in the Pathogen	
Genes in the Plant	**AVR / AVR or AVR / avr**	**avr / avr**
R / R or R / r	Disease Resistance	Disease Susceptibility
r / r	Disease Susceptibility	Disease Susceptibility

Figure 1.1 Illustration of the gene-for-gene concept.

Disease resistance is most often characterized by an incompatible reaction known as the hypersensitive response (reviewed in Dangl *et al.*, 1996). A series of defense responses in the plant lead to localized necrosis at the site of infection, which prevents further spread of the pathogen. The hypersensitive response is thought to be down-regulated in normal, non-infected plants and induced upon an incompatible disease reaction (Greenberg *et al.*, 1994). As reviewed in Flor (1971), the flax/rust gene-for-gene model is not an exception, and is applicable to a large number of other plant/pathogen combinations.

1.1.3.1 Pathogen Avirulence Genes

The first avirulence gene (*avrA*) was cloned from the bacterium *Pseudomonas syringae* pv. *glycinea* race 6, which normally infects soybean plants (Staskawicz *et al.*, 1984). It was found that, when *avrA* was transferred to three other races of the pathogen, it conferred the same host specificity as race 6 in each case. As new avirulence genes were cloned from other pathogens, it was also shown that they could be transferred to

related pathogens and affect their host range. For example, *P.s.* pv. *tomato* can be induced to cause an incompatibility reaction in some soybean varieties if it receives the appropriate avirulence genes from *P.s.* pv. *glycinea* (Kobayashi *et al.*, 1989). These results confirm one half of the gene-for-gene theory, as they strongly suggest that a single gene in the pathogen is sufficient to cause avirulence.

Many avirulence genes have been sequenced to this date, but their function remains largely unknown. The very presence of avirulence genes in pathogens is rather puzzling. Why would pathogens produce proteins that allow their hosts to neutralize them? Perhaps the simplest explanation is that these genes code for proteins involved in pathogenesis, and that the cellular machinery of their host plants learned to recognize them as a sign of infection (Jackson and Taylor, 1996). The outcome of host/pathogen interactions would therefore be determined by an evolutionary race for the production of different avirulence genes in the pathogen and the corresponding resistance gene specificities in the plant.

Recombination is presumed to be important in the generation of new specificities in both the pathogen and the plant. Yang and Gabriel (1995) showed that intragenic recombination in a bacterial pathogenicity/avirulence gene was indeed responsible for the alteration of host specificity. The gene *pthA* from *Xanthomonas citri* is required for virulence on citrus plants but can also act as an avirulence gene in hosts that possess the appropriate resistance gene. Yang and Gabriel (1995) found that, upon intragenic recombination among tandem repeats in the central portion of *pthA*, a number of different, non-essential genes could be generated, with a gain or loss of avirulence on bean, citrus, and cotton plants.

Mode's mathematical model of gene-for-gene interactions predicts that a system where the avirulence genes are inherited independently from each other, and the resistance genes are allelic, as is most often the case, would insure the long term survival of both the pathogen and the host (reviewed in Flor, 1971).

1.1.3.2 Plant Disease Resistance Genes

The *Pto* gene was the first cloned plant disease resistance gene to conform to the gene-for-gene theory. It was cloned from tomato in 1993 using a map-based approach, and was shown to confer resistance to *Pseudomonas syringae* pv. *tomato* in complementation experiments (Martin *et al.*, 1993). The *Hm1* gene of maize had been cloned in 1992 (Johal and Briggs, 1992). However it confers toxin resistance and, as such, does not fit the gene-for-gene theory and belongs to a class of its own. Since then, 12 other resistance genes (R genes) have been isolated in both mono- and dicotyledonous plants (Table 1.1). All R genes can be grouped into 5 classes according to the predicted structure of their gene product (Baker *et al.*, 1997). It is likely however that, as new R genes are cloned, and new structures are described, new classes will have to be formed. For example, a new R gene conferring resistance to wheat rust has recently been identified, and it encodes a serine/threonine kinase with a new type of extracellular domain (Feuillet *et al.*, 1997).

Perhaps the most surprising characteristic of plant R genes is their striking structural similarity, despite the fact that they confer resistance to a wide variety of pathogens, from viruses (Whitham *et al.*, 1994) to nematodes (Cai *et al.*, 1997). The most

common motif is a series of imperfect leucine-rich repeats (LRR) thought to be involved in protein-protein interactions (Kobe and Deisenhofer, 1993). This motif is found in 12 of

Table 1.1: Cloned plant disease resistance genes (adapted from Baker *et al.*, 1997)

Class	R gene	Plant	Pathogen	Structure[a]	Reference[b]
1	*RPS2*	Arabidopsis	*Pseudomonas syringae* pv. *tomato*	LZ-NBS-LRR	(1), (2)
	RPM1	Arabidopsis	*P. syringae* pv. *maculicola*	LZ-NBS-LRR	(3)
	Prf	Tomato	*P. syringae* pv. *tomato*	LZ-NBS-LRR	(4)
	N	Tobacco	Tobacco Mosaic Virus (TMV)	TIR-NBS-LRR	(5)
	L^6	Flax	*Melamspora lini*	TIR-NBS-LRR	(6)
	M	Flax	*M. lini*	TIR-NBS-LRR	(7)
	RPP5	Arabidopsis	*Peronospora parasitica*	TIR-NBS-LRR	(8)
	I2C-1	Tomato	*Fusarium oxysporum*	NBS-LRR	(9)
2	*Pto*	Tomato	*P. syringae* pv. *tomato*	Protein Kinase	(10)
3	*Cf-9*	Tomato	*Cladosporium fulvum*	LRR-TM	(11)
	Cf-2	Tomato	*C. fulvum*	LRR-TM	(12)
	$HS1^{pro-1}$	Sugar Beet	*Heterodera schachtii*	LRR-TM	(13)
4	*Xa21*	Rice	*Xanthomonas oryzae* pv. *oryzae*	LRR, Protein Kinase	(14)
5	*Hm1*	Maize	*Cochliobolus carbonum* race 1	Toxin Reductase	(15)

[a] Structure refers to the protein domains predicted from the amino acid sequence: leucine zipper (LZ), nucleotide binding site (NBS), leucine-rich repeats (LRR), Toll-IL-1R homology region (TIR), and transmembrane domain (TM).

[b] References: (1) Bent *et al.* (1994), (2) Mindrinos *et al.* (1994), (3) Grant *et al.* (1995), (4) Salmeron *et al.* (1996), (5) Whitham *et al.* (1994), (6) Lawrence *et al.* (1995), (7) Anderson *et al.* (1997), (8) Parker *et al.* (1997), (9) Ori *et al.* (1997), (10) Martin *et al.* (1993), (11) Jones *et al.* (1994), (12) Dixon *et al.* (1996), (13) Cai *et al.* (1997), (14) Song *et al.* (1995), and (15) Johal and Briggs (1992).

the 14 resistance genes isolated so far. The 8 genes belonging to class 1, the largest class of R genes so far, also possess a nucleotide binding site (NBS). This similarity suggests an evolutionary conserved mechanism for the detection of invading organisms.

In addition, the plant disease response is also very similar among distantly related plants, regardless of the pathogen. The most common scenario involves the hypersensitive response described above. Upon recognition of the pathogen, the resistance genes are thought to trigger a signal transduction cascade which leads to an

enhanced expression of pathogenesis related (PR) genes, an oxidative burst which is characterized by an accumulation of reactive oxygen species which act both as antimicrobial agents and as cell wall protein crosslinking agents (Levine *et al.*, 1994), and to localized cell death. Chandra *et al.* (1996) showed that the Pto protein is necessary for the production of a resistance-specific oxidative burst. Zhou *et al.* (1997) have also shown, using the two-hybrid system, that Pto can bind transcriptional activators of several PR genes. Finally, activation of local disease response mechanisms leads to the activation throughout the plant of a broad-spectrum resistance known as systemic acquired resistance (SAR) (reviewed in Ryals *et al.*, 1996).

As with avirulence genes, plant R genes can be transferred to other plants and confer new resistance specificities to the host plant. For example, it was shown that *Pto* can be transferred from tomato to tobacco plants and confer resistance to *P. syringae* pv. *tabaci* expressing *avrPto* (Thilmony *et al.*, 1995; Rommens *et al.*, 1995). This suggests that a Pto-like signal transduction pathway is conserved in the Solanaceae. There is no formal evidence however that R gene-mediated signal transduction cascades are conserved across plant families, and that R genes can be transferred at will. We can speculate that this is the case however, because of the sequence conservation among R genes in very different plant families, and because of the fact that the hypersensitive response occurs in mono- and dicotyledonous species alike.

Nonetheless, even if a resistance gene can function in a distant plant family, it is useless unless the pathogen to which it confers resistance is able to infect the new host plant (Moffat, 1994). The engineering of plants with new disease resistance specificities is therefore more likely to result from the transfer of R genes among closely related plant

species, or at least species that are affected by the same pathogens. In addition, once we understand the basis for specificity, it should be possible to modify existing resistance genes as to produce new recognition capabilities.

In human gonadotropins receptors, LRRs are the domains responsible for ligand binding specificity (Braun *et al.*, 1991). Analysis of the flax rust resistance gene *M* reveals that the LRR region may also be associated with disease resistance specificity. Mutant alleles of the *M* gene had lost one of their leucine-rich repeats, presumably through unequal recombination, and had also lost the ability to confer resistance to flax rust (*Melampsora lini*) (Anderson *et al.*, 1997). There was no evidence however that new specificities were generated in these mutants, though it is possible that the avirulence genes corresponding to these new alleles have not yet been uncovered. It was found that the *M* gene lies within a cluster of approximately 15 similar genes but that none of the other genes confer the *M*-specific resistance (Anderson *et al.*, 1997). The *Cf-2* (Dixon *et al.*, 1996), *Cf-9* (Jones *et al.*, 1994), *N* (Whitham *et al.*, 1994), *Pto* (Martin *et al.*, 1993), and *Prf* (Salmeron *et al.*, 1996) genes were also found to be part of gene clusters. In fact, *Prf* and *Pto* are part of the same cluster, along with the *Fen* gene, which encodes sensitivity to the organophosphorous insecticide fenthion (Loh and Martin, 1995). The *Pto* and *Fen* genes share 80% identity and both encode functional protein kinases (Loh and Martin, 1995). They are thought to be part of the same signal transduction cascade, as they both trigger an hypersensitive response upon exposure to their respective stimuli (Loh and Martin, 1995). Jia *et al.* (1997) recently showed, however, that the kinase function of *Pto* and *Fen* was not sufficient to confer their respective phenotypes in tomato plants, as new alleles were found that encode functional kinases but are deficient in

interacting with other members of the *Pto/Fen* signal transduction cascade and in eliciting the hypersensitive response. These new alleles should be helpful in dissecting the structure/function relationship in plant disease resistance genes.

It is conceivable that the *Fen* gene was generated by recombination involving *Pto*, and that, instead of gaining a new disease resistance specificity, it gained the ability to recognize a different signal. The way by which a living cell recognizes and responds to very different external stimuli may therefore evolve from a limited number of mechanisms. Indeed, conserved motifs that seem to play a key role in disease resistance, such as LRRs, are found in a variety of other proteins (Kobe and Deisenhofer, 1993). For example, the *Prf* gene is a class 1 R gene and shares no homology with *Pto* (Salmeron *et al.*, 1996). Rather, it seems to complement *Pto* and *Fen* in their respective functions, as it is required for resistance to *P. syringae*, as well as for sensitivity to fenthion (Salmeron *et al.*, 1996). The arrangement of similar R genes in clusters suggests homologous recombination as a possible mechanism for the production of new resistance genes, as is the case with avirulence genes. This has been shown, in fact, in studies of the maize rust resistance gene family *Rp1*, where recombination was found to be responsible, not only for R gene instability (Sudupak *et al.*, 1993), but also for the generation of new disease resistance specificities that were not found in either parent (Richter *et al.*, 1995). The presence of *Prf* within the *Pto/Fen* cluster, however, also shows that R genes can be clustered by function as well as by sequence homology.

Not all R genes however are part of clusters. Some resistance specificities are encoded by multiple alleles at a single locus. The L^6 gene, for example, hybridizes to a large number of similar genes, but they all map to the *M* locus (Ellis *et al.*, 1995). The

different known alleles of L^6 all map exactly to the same locus, and it is thought that the different L^6 resistant specificities are all alleles of the same gene (Lawrence *et al.*, 1995). Interestingly, a mutant flax plant that possesses a recombination event within the L^6 locus exhibited partial resistance to the rust fungus (Lawrence *et al.*, 1995). Thus both intra and intergenic recombination are presumably involved in the evolution of plant disease resistance gene specificities.

The *RPS2* (Bent *et al.*, 1994; Mindrinos *et al.*, 1994) and *RPM1* (Grant *et al.*, 1995) genes of *Arabidopsis* also occur as single R genes at their respective loci. Mutants at the *RPS2* locus correspond to various alleles of the same gene. Most of these correspond to point mutations that alter the amino acid sequence of the gene. Point mutations in *RPM1* were also identified but, surprisingly, naturally occurring susceptible plants lack the gene altogether (Grant *et al.*, 1995). An interesting feature of *RPM1* is that it expresses specificity for two different avirulence proteins of *P. syringae*. The basis for this dual specificity is not known. Grant *et al.* (1995) suggest that plants may evolve R genes with multiple specificities to minimize the number of R genes required to cope with the plethora of pathogens that they must face.

1.1.3.3 Modern Evidence

Until 1997, the molecular basis for the gene-for-gene theory was unknown. One of the favored hypothesis was that the R gene products act as specific receptors for the pathogen elicitors, which are encoded directly or indirectly by the corresponding avirulence genes (Keen, 1990). This posed a conceptual problem however, as R genes were being cloned and sequenced, for most of them, except for the class 3 genes, were

thought to code for intracellular proteins (reviewed in Baker *et al.*, 1997). Recent findings have helped to alleviate this problem, as it was shown that molecular recognition of many avirulence proteins occurs inside plant cells (Gopalan *et al.*, 1996; Leister R.T. *et al.*, 1996; Van den Ackerveken *et al.*, 1996). In transient and stable plant transformation experiments, avirulence proteins can elicit a resistance specific hypersensitive response.

It was also found that *Escherichia coli* can be induced to cause an hypersensitive response in *A. thaliana* expressing *RPM1* if the bacteria express the *AvrB* gene and the *P. syringae* hypersensitive response and pathogenesis (Hrp) type III protein secretion system (Gopalan *et al.*, 1996). Bacteria expressing *AvrB* or the *Hrp* gene complex alone do not elicit an hypersensitive response (Gopalan *et al.*, 1996). Phytopathogenic Pseudomonad and Xanthomonad bacteria, like the animal pathogens *Yersinia*, *Salmonella*, and *Shigella* (Rosqvist *et al.*, 1995), transfer their virulence factors across the host cell's plasma membrane using a type III protein secretion system (reviewed in Alfano and Collmer, 1997). These facts taken together suggest that at least some bacterial avirulence genes have a role in virulence, which would explain why they persist even once the host has evolved a corresponding resistance gene.

Some plant R genes do possess a predicted transmembrane domain (Table 1.1), suggesting that they respond to external signals. For example, the class 3 genes *Cf-9*, *Cf-2*, and $HS1^{pro-1}$ appear to have an extracellular LRR domain, a membrane anchor, and a small cytoplasmic tail (Parker and Coleman, 1997). Supposing that LRR domains are responsible for interacting with avirulence proteins, these genes would be well positioned to detect the avirulence proteins from pathogens, such as *C. fulvum*, that colonize the intercellular spaces (Parker and Coleman, 1997). In agreement with this hypothesis,

targeting to the apoplast of the Avr9 peptide from *C. fulvum* in transgenic tomato plants triggers programmed cell death (Hammond-Kosack *et al.*, 1994). In short, it would seem that there is a strong correlation between the localization of resistance proteins and the location of avirulence proteins during infection. This gives strong evidence for the direct physical interaction of resistance proteins and avirulence gene products.

This hypothesis was finally proven formally fifty years after Flor's experiments. In two elegant studies using the yeast two-hybrid system, Scofield *et al.* (1996) and Tang *et al.* (1996) showed that AvrPto interacts with the Pto kinase, and that mutations in either gene which disrupt this interaction also suppress the disease resistance response in tomato plants. These findings are somewhat surprising in that Pto does not possess a LRR domain, thought to be responsible for recognition of avirulence proteins. It does require a LRR protein, Prf, however, for resistance to take place (Salmeron *et al.*, 1996). Scofield *et al.* (1996) suggest that Prf perhaps recognizes the Pto/AvrPto complex, and that it will be interesting to see whether disease resistance always requires a protein kinase to recognize the avirulence protein, or if the *Pto* system is unique in this way.

As mentioned previously, the function of avirulence genes is still largely unknown and, except for the toxin reductase *Hm1*, and the protein kinases *Pto* and *Xa21*, the exact function of R genes remains unclear as well. Yet, our increasing knowledge of the pathogen virulence and plant resistance systems is allowing us to dissect the disease response and to gain a better understanding of the complex interactions between plant pathogens and their hosts.

1.1.3.4 Exceptions

Not all plant disease resistance responses conform to the gene-for-gene theory. The maize toxin reductase gene *HM1* is an example of a single gene that confers specific resistance to strains of the fungus *C. carbonum* that produce the HC-toxin, but doesn't suit the gene-for-gene model, because interaction with an avirulence gene is not involved (Johal and Briggs, 1992). R gene mediated resistance, though a strong and specific defense mechanism, only represents a small subset of all the weapons that a plant possesses to defend itself from invading pathogens.

Plants possess a myriad of more general defense responses. A complete review of all these mechanisms is outside the scope of this thesis. It is important however to recognize that plant disease resistance results from the concerted action of several lines of defense.

The first line of defense consists of preformed physical and chemical barriers, such as the waxy cuticle layer and the cell wall (especially of epidermal cells) (Keen, 1992). A water-repellent layer such as the cuticle on plant surfaces prevents fungal spores from germinating and bacteria from multiplying (Agrios, 1988). Thick cell walls also help to slow down the spread of pathogens within the plant (Agrios, 1988).

Plants also possess a vast array of antimicrobial chemical compounds which may be important in defense. The exact role of preformed chemical defenses is still not clear, though at least one compound, avenacin A-1, a saponin produced by oat plants, seems to play a key role in resistance to the "take-all" fungal disease. Presence of an avenacin detoxifying enzyme in the fungus was shown to be essential for infection (reviewed in Osbourn, 1996).

When a pathogen penetrates the first line of plant defenses, it usually induces a number of other responses at the tissue, cellular, and molecular level in the plant. The most common examples of inducible defenses include the formation of a cork layer, cutting off the supply of water and nutrients to the site of infection, the deposition of gums in and around the infected cells, thickening of the plant cell wall, and production of phenolic compounds such as the phytoalexins, which are usually produced upon injury (Agrios, 1988). This is by no means an exhaustive list, and many defense mechanisms have evolved in specific host/pathogen combinations and not in others (Agrios, 1988). In short, though R genes participate in specific recognition of invading pathogens, plants are able to recognize and respond to general injury.

1.1.3.5 Homology to Animal Systems

Though plants do not possess the type of immune system found in animals, some interesting analogies can be drawn between components of plant and animal disease resistance mechanisms. For example, the class 1 genes *N*, *L6*, *M*, and *RPP5* all share homology with the *Drosophila melanogaster Toll* gene, which encodes a receptor involved in establishing the dorsal/ventral patterning in the embryo, and with the mammalian interleukin-1 receptor (Il-1R) (Staskawicz et al., 1995) (see Figure 1.2). Interestingly, the Toll receptor and its signal transduction cascade are involved in *Drosophila* immunity later in development (reviewed in Belvin and Anderson, 1996). The Il-1R pathway is also involved in innate immunity, leading to a rapid, non-specific disease response (reviewed in Belvin and Anderson, 1996).

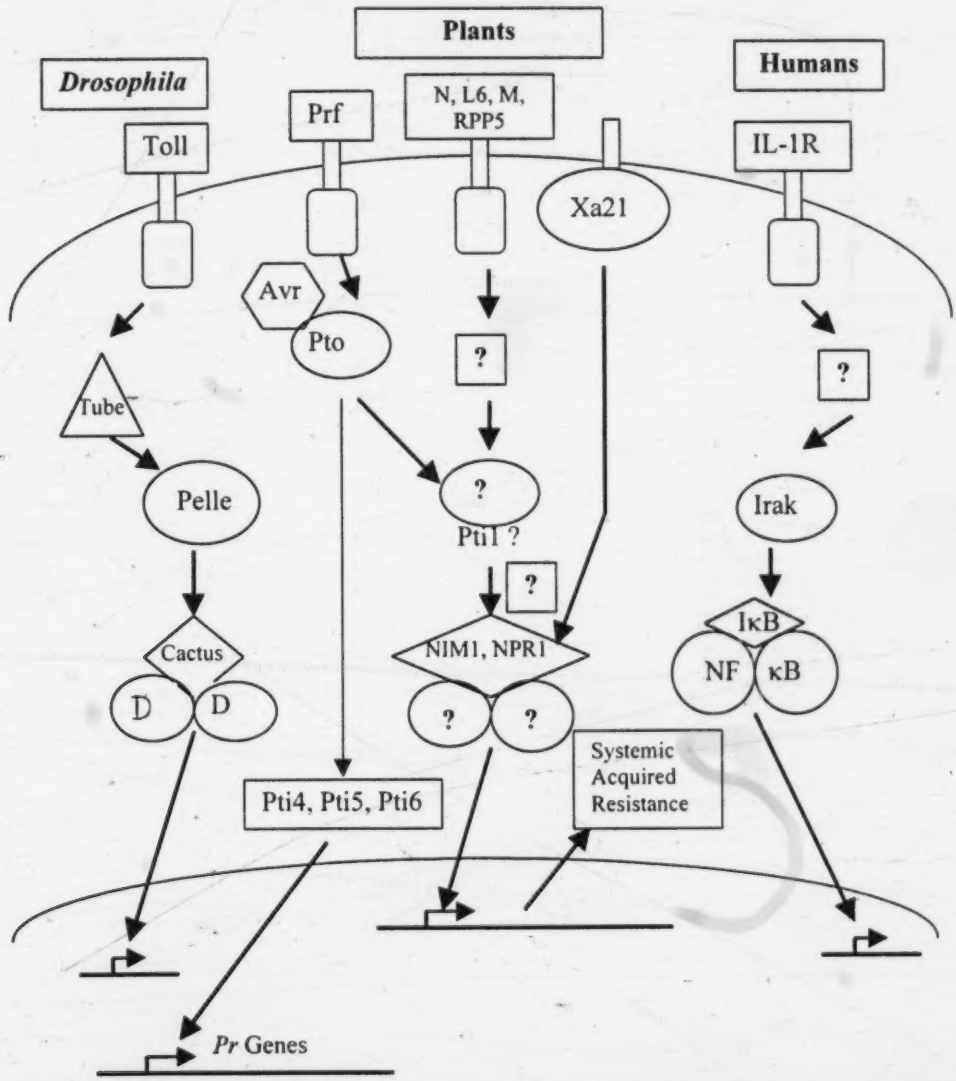


Figure 1.2. Similarities between gene-for-gene plant disease resistance, the *Drosophila* Toll pathway, and the human IL-1R pathway. Similarly shaped proteins indicate similarities between the signal transduction components involved in plant and animal immunity. The *Drosophila* Toll pathway is indicated on the left, while the human IL-1R pathway is on the right. The most striking similarities are between the Toll/IL-1R receptors and some plant R proteins, the Pelle/Irak serine/threonine kinases and the plant Pto, Pti1, and Xa21 proteins, and between the Cactus/IκB transcription factor inhibitors and the *Arabidopsis* NIM1 and NPR1 proteins (see text for details).

However, the homology between the plant and animal systems does not stop there. Signal transduction components downstream of Toll and IL-1R are similar to a number of plant genes. The Pelle and Irak cytoplasmic serine/threonine kinases, for example, are similar to the plant R proteins Xa21 and Pto (Baker *et al.*, 1997), as well as to Pti1 and Fen (Belvin and Anderson, 1996). Pti1 is a cytoplasmic kinase which is phosphorylated by Pto (but not by Fen) and is involved in the hypersensitive response (Zhou *et al.*, 1995). Interestingly, two transmembrane protein kinases with extracellular LRR domains, CLAVATA1 (Clark, 1997) and ERECTA (Torii *et al.*, 1996), which are involved in plant morphogenesis, also belong to this kinase family (Baker *et al.*, 1997). Belvin and Anderson (1996) refer to this group as the serine/threonine innate immunity kinases (SIIK). Whereas the Toll pathway plays a role both in development and the immune response, the plant developmental and immune responses seem to be separate.

Further downstream of the Toll and IL-1R pathways are the Cactus and IκB ankyrin repeat proteins which are involved in binding transcription factors in the cytoplasm, thus preventing the latter to enter the nucleus (reviewed in Belvin and Anderson, 1996). At this point, it is not surprising to realize that the *Arabidopsis NIM1* (Ryals et al., 1997) and *NPR1* (Cao et al., 1997) genes (which were found to be alleles of the same gene) also possess ankyrin repeats, are homologous to IκB and are involved in positively regulating SAR. It is therefore likely that NIM1 and NPR1 function in a similar fashion, inhibiting transcription factors of PR genes until a disease signal is perceived by the cell.

Moreover, salycilic acid (SA) and reactive oxygen species, such as H_2O_2, which are known to be involved in local disease resistance, and in the induction of SAR in plants (White, 1979; Levine *et al.*, 1994), are also involved in the human IL-1R defense pathway (Kopp an Ghosh, 1994; Schreck *et al.*, 1991). Surprisingly however, SA has an inhibitory effect on NF-κB-mediated transcription in animals (Kopp and Ghosh, 1994), whereas it induces the production of defense proteins in plants (Delaney *et al.*, 1994; Shah and Klessig, 1996). Hydrogen peroxide, on the other hand, strongly induces defense responses in both plants (Levine *et al.*, 1994) and animals (Schreck *et al.*, 1991; Meyer *et al.*, 1993).

The relationship between H_2O_2 and SA in plant disease resistance is still unclear, and is the subject of a strong controversy. Whereas some believe that SA induces H_2O_2 to accumulate in the cell by inactivating catalases and peroxidases (Chen *et al.*, 1993; Durner and Klessig, 1995, 1996), others believe that hydrogen peroxide levels accumulate via a SA-independent mechanism in vivo (Bi *et al.*, 1995; Kvaratskhelia *et al.*, 1997). This controversy appears to be due to the difficulty of working with catalases and active oxygen species. Also, as Durner and Klessig (1996) suggest, SA might influence catalases differently in infected and healthy cells, as its action seems to depend on the redox status of the cell. In any case, all agree that SA and H_2O_2 play an important role in the plant disease response, and a hint as to their exact function may come from the study of animal systems. In this respect, it is interesting to note that H_2O_2 in animals mediates the release of arachidonic acid into the cell, and that this response activates protein kinase C (PKC), which stimulates the activation of transcription factors (Rao *et al.*, 1993). Arachidonic acid release is also involved in plant PR gene activation (Chen *et*

al., 1993). Moreover, phosphorylation cascades play a key role in the plant defense response (Després *et al.*, 1995), and Subramaniam *et al.* (1997) have recently shown that a plant PKC homolog is involved in the transcriptional activation of the pathogenesis-related gene *PR-10a* in potatoes.

The similarities in disease resistance in plant and animals extend to the mechanisms used by pathogens to infect their host. As mentioned before, it is thought that some bacterial plant pathogens use type III secretion systems to inject virulence factors into the cytoplasm of their host. This mechanism is used by several animal pathogens, including *Yersinia*, *Shigella*, and *Salmonella* (Rosqvist *et al.*, 1995). Furthermore, it was found that the *Pseudomonas aeruginosa* strain UCBPP-PA14 can infect mice, as well as specific ecotypes of *A. thaliana*, using the same virulence factors (Rahme *et al.*, 1995).

Together, these lines of evidence strongly suggest that at least some modes of infection by plant and animal pathogens may be conserved, and that the means by which a host responds to infection might be ancestral to the divergence of the plant and animal kingdoms, more than one billion years ago (Belvin and Anderson, 1996).

1.2 Isolating Plant Disease Resistance Genes

With an exponentially increasing world population, and a steady decrease in total arable land, it becomes increasingly important to develop sound, sustainable increases in plant yields, as well as efficient post-harvest conservation methods. Gibson and Somerville (1993) state that, in the case of well-characterized systems, crops may be

rationally improved via transgenic technology. This is currently limited, however by a definite lack of cloned genes.

Until 1992, this was also true of plant R genes. Over the last five years however, a myriad of plant disease resistance genes and their homologues have been isolated. This dramatically changed the study of plant disease resistance and will help us considerably in generating crops with new and improved disease resistance specificities.

In more complex systems, such as systemic acquired resistance, it may be more efficient to use molecular markers, closely linked to the different genes involved in the pathway, to assist the introgression of these genes into the cultivars of interest (Allen, 1994). Availability of molecular markers will also help in the cloning of closely linked genes of interest.

Several methods are available to isolate plant genes (reviewed in Gibson and Somerville, 1993). Transposon tagging and positional cloning, the two methods used in the isolation of plant R genes so far, as well as the isolation of plant R gene homologues using the polymerase chain reaction (PCR), hybridization, and random DNA sequencing, are discussed below.

1.2.1 Transposon tagging

This method is based on the identification of mutants created by the insertion of transposable elements within a gene of interest, or within its promoter. The gene can then be identified by hybridizing a genomic library with a probe corresponding to the transposon (Fedoroff *et al.*, 1984).

The properties of Barbara McClintock's maize *Activator* (*Ac*) transposable element have been used in cloning the plant R genes *N*, *Cf-9*, L^6, and *M*. The maize *Hm1* gene, on the other hand, was cloned using the *Mutator* (*Mu*) element, which was first described at the molecular level by Barker *et al.* (1984).

The *Ac* element was cloned in 1983 (Fedoroff *et al.*, 1983), and then used for the first time to clone a plant gene, the maize *bronze* gene, in 1984 (Fedoroff *et al.*, 1984). A genomic library from mutants at the *bronze* locus was probed with the internal region of the *Ac* element. Positive clones were used to probe genomic DNA of plants with known genotypes. Clones that co-segregated with the *bronze* gene were identified. Fedoroff *et al.* (1984) suggested that this method could be widely used in the isolation of plant genes, but until recently, the use of *Ac* elements was restricted to maize. It wasn't until 1993, that an *Ac* derivative was used to clone a gene from a dicotyledonous plant species (Chuck *et al.*, 1993).

Whitham *et al.* (1994) used a slight modification of Fedoroff's protocol in their cloning of the tobacco *N* gene. They used inverse PCR (iPCR) (Ochman *et al.*, 1988) to identify the sequences flanking the inserted *Ac* element. Genomic DNA from the mutant line was digested with an enzyme that doesn't cut within *Ac*, ligated to form circular DNA molecules, and linearized again with an enzyme that cuts once within *Ac*. PCR was then performed using primers specific for the ends of *Ac*. The resulting PCR products were used as probes to identify genomic and cDNA clones corresponding to the *N* gene.

Though transposon tagging has proven very useful in the isolation of numerous genes, it is limited to the species that possess functional transposable elements. Furthermore, a mutation must be identified before the gene of interest can be isolated,

and genes with no easily observable phenotypes often must be cloned using other approaches.

1.2.2 Genetic Mapping and Map-Based Cloning

Genetic mapping is based on the fact that, upon meiosis, chromosomes are able to exchange parts of their DNA, thus creating new gene combinations in the resulting gametes. The probability that such an exchange, or recombination event, will occur between two genes increases as the physical distance separating the two genes increases. Genetic distance is therefore based on the recombination frequency between two DNA segments, or markers, and is measured in centimorgans (cM), which are roughly equivalent to 1% recombination. There is no linear relationship however, between physical and genetic distance, as some chromosomal regions are more prone to recombination than others. This relationship between genetic and physical distance therefore needs to be determined for every chromosomal region of interest.

Allen (1994) discusses the utility of genetic maps by comparing them to roadmaps. The best way to access a region of interest is to use a complete, accurate map. Prior to the 1980s, the development of such maps was hindered by the relative scarcity of markers available. The markers used were mainly based on traceable phenotypes, which can be influenced by the environment to a large extent. Also, visible phenotypes correspond to a very small subset of all possible mutations that can be mapped. The advent of molecular markers in the 1980s and 90s, such as restriction fragment length polymorphisms (RFLP) (Botstein *et al.*, 1980), random amplified polymorphic DNA

(RAPD) (Williams *et al*., 1990), and microsatellite (Akkaya *et al*., 1992) markers, has promoted a dramatic increase in the saturation and accuracy of genetic maps.

The advantage of DNA markers over phenotypic markers lies in the fact that they are not, in theory, affected by the environment (though methylation effects can affect the scoring of some molecular markers – Derek Lydiate, personal communication), and that they need not correspond to a visible trait. Many DNA markers are also amenable to automation, thus increasing considerably the potential number of traits that can be scored in a given amount of time (Rafalski and Tingey, 1993).

DNA markers closely linked to a trait of interest can thus be used to follow the segregation of that trait in a population with much higher accuracy. This allows for maker-assisted selection (MAS) in the process of crop improvement. MAS not only saves time and resources by improving selection accuracy, it also allows for the selection of several traits simultaneously (Allen, 1994).

Over the last ten thousand years, plant breeding has resulted in highly homogenous, high performance plant varieties. Though the Green Revolution of the 1940s-60s has allowed a dramatic increase in crop yields, global grain production seems to have reached a plateau (Mann, 1997), partly because of the lack of residual variation in the crops in use (Tanksley and McCouch, 1997). This genetic uniformity not only hinders the progress of plant breeding, but it leaves the crops vulnerable to severe pathogen and insect epidemics, as exemplified by the U.S. corn leaf blight disaster of the 1970s (Tanksley and McCouch, 1997). Fortunately, the need to preserve wild, or "exotic" germplasm was recognized, and reinforced after the leaf blight epidemic, and over 700 gene banks have been established worldwide (Tanksley and McCouch, 1997).

Nonetheless, traditional plant breeding relies on the rejection of all but the favorable traits and less productive wild germplasm is often excluded from breeding programs (Mann, 1997). Progresses made during the Green Revolution however, were made by introgressing dwarfing genes from wild varieties of wheat and rice into modern cultivars (Mann, 1997). More recently, Tanksley and McCouch (1997) have developed a method, using highly saturated molecular maps, for utilizing the valuable variation present in primitive, apparently low producing wild varieties. They achieved a 48% yield increase in tomato, using this method, when the average for this crop is of less than 1% annually. These results show that genetic maps can be used to trace the genomes of wild varieties in the crosses with elite varieties, thus identifying which of the "wild" alleles confer an advantage over the elite parent. This suggests that, at least with crops for which saturated genetic maps are available, it should be possible to achieve considerable crop improvements by harvesting the variation stored in gene banks and using it to widen the genetic base of our elite crops (Tanksley and McCouch, 1997).

This applies to the introgression of new disease resistance genes, especially as we are starting to identify specific chromosomal regions associated with R-gene clusters. There is considerable interest in the pyramiding of genes conferring resistance to the same pathogen, as this should increase the "life span" of each R gene. Using molecular markers closely associated with R gene clusters should allow the simultaneous introgression of numerous resistance specificities.

Genetic maps can also be used for the actual cloning of genes for which mutations are known. This is known as map-based cloning (reviewed in Gibson and Somerville,

1993). This approach has been used in the cloning of the plant R genes *Pto*, *RPS2*, *RPM1*, *Xa21*, *Cf-2*, *I2C-1*, *Prf*, *RPP5*, and $Hs1^{pro-1}$.

Map-based cloning first involves the identification of closely linked markers on both sides of the gene of interest. As Tanksley *et al.* (1995) mention, this was not always the case. Previous efforts used to focus on the isolation of large, overlapping DNA clones encompassing the region containing the gene of interest, as the isolation of large numbers of molecular markers was more difficult than it is today. The isolation of genes in this way ("chromosome walking") was therefore very difficult and time-consuming, especially for plants with very large genomes and large stretches of repetitive DNA (Tanksley *et al.*, 1995). Since the development of high throughput DNA marker technologies such as RAPDs and of strategies that allow the quick identification of markers in a defined region, the focus in map-based cloning has shifted towards the identification of very closely linked markers first. This is followed by the identification of a single clone spanning the region of interest and is known as "chromosome landing" (Tanksley *et al.*, 1995).

Cloning of the *Pto* gene (Martin *et al.*, 1993) is an example of chromosome landing (Tanksley *et al.*, 1995). Mapping of *Pto* was accomplished by screening 251 F_2 plants derived from a cross between homozygous resistant and susceptible parents. Using a marker that cosegregated with *Pto*, they isolated a yeast artificial chromosome (YAC) clone containing a 400 kb DNA fragment from tomato and showed that it spanned the *Pto* locus. This clone was then used to probe a tomato leaf complementary DNA (cDNA) library. A cDNA clone was identified that cosegragated with *Pto* when mapped in tomato. When they rehybridized the cDNA library with the candidate clone, however,

they found at least 6 new cDNAs, falling into two size classes, with homology to the first one. To find out which clone corresponded to Pto, they transformed susceptible tomato plants with a representative cDNA from both size classes. Using this approach, they found that a 2.4 kb cDNA could confer resistance to pathogens producing *AvrPto*.

1.2.2.1 High Resolution Plant Genetic Mapping, Pooled Sampling, and Bulked Segregant Analysis

Saturated genetic maps provide a powerful tool for the study and isolation of genes. However, though modern technology allows for the rapid screening of hundreds of molecular markers per day, genetic mapping of entire genomes remains an intimidating task, due to the large populations that need to be analyzed, and to the high number of markers that must be mapped. For example, it is estimated that, to have a 90% chance to find at least one marker within 200 kb of a target sequence in tomato (which has a genome size of 1 billion base pairs), approximately 6000 markers would have to be analyzed (Tanksley *et al*., 1995). Similarly, to have a 95% chance to identify a plant with a single recombination event within 0.1 cM of a gene of interest would require that more than 1500 diploid plants (3000 meioses) be analyzed (Tanksley *et al*., 1995). Because of this, saturated maps are still not available for many important crops. This prompted the development of strategies that allow for the quick identification of individuals that possess recombination events in the chromosomal region of interest only, as well as of markers which lie in that region. Three of these methods, pooled-sampling, near-isogenic lines (NILs), and bulked segregant analysis, and their application to the mapping of plant disease resistance genes, are briefly described below.

Plant near-isogenic lines are generated by crossing two parents which differ at a locus of interest. The F_1 hybrid is then backcrossed to the recipient parent. From the progeny, a plant containing the dominant allele of the target gene is selected and backcrossed to the recipient parent once more. This backcrossing step is repeated until the 7^{th} or 8^{th} generation is obtained. Once again, a plant having the dominant allele is selected, selfed and, from the progeny, a plant homozygous for the locus of interest is selected. This plant is said to be isogenic to the recipient parent, as it should theoretically differ only in a small region surrounding the gene of interest (reviewed in Tanksley *et al.*, 1995). DNA from both lines is extracted and subjected to molecular marker analysis. Such plant lines are very useful, as markers that give a polymorphism between two different NILs should theoretically be closely linked to the gene of interest. This method has been used in the identification of markers closely linked to the disease resistance gene *Pto* (Martin *et al.*, 1991).

All types of markers can be used to map target chromosomal regions with NILs, though RAPDs and RFLPs are the most widely used. RFLPs are detected by probing genomic DNA, which has been digested with a restriction enzyme, with either an arbitrary genomic or cDNA clone, or a known sequence. The polymorphisms are scored as different band sizes on Southern blots. RFLPs are codominant in that they can differentiate homozygotes from heterozygotes at a particular allele. This is in contrast to RAPDs, which are scored as the presence or absence of a PCR product when genomic DNA is amplified with short, arbitrary primers (Williams *et al.*, 1990; reviewed in Rafalski and Tingey, 1993).

Bulked segregant analysis (Michelmore *et al.*, 1991; Giovannoni *et al.*, 1991) is based on the same theoretical principle as NILs, except that the F_2 plants derived from the cross described above are grouped into plants that are homozygous dominant, heterozygous, and homozygous recessive at the target locus. The genotyping of F_2 individuals is usually achieved by analyzing the segregation patterns of the gene of interest in the F_3 generation derived from each F_2 plant. The heterozygous plants are discarded, and the two remaining groups, or bulks, are used in the same way as NILs. The two bulks should differ only at the target locus for, in theory, if enough individuals are included in each bulk, all possible alleles at every other locus in the genome are represented in both bulks (Michelmore *et al.*, 1991; Giovannoni *et al.*, 1991). Markers unlinked to the gene of interest should therefore produce no polymorphisms between the two bulks. The number of individuals in each bulk dictates in part the maximum genetic distance at which a given marker can lie from the target locus and still produce a polymorphism (Williams *et al.*, 1993). With a theoretical infinite number of individuals in each bulk, every polymorphism found would lie within the gene of interest. In practice however, some markers are more difficult to score than others, and there is therefore a limit in the ability to detect rare alleles in a DNA pool, thus creating a minimum detection window size around the target locus (Williams *et al.*, 1993).

Bulked segregant analysis has several advantages over NILs. First, bulks can be generated as soon as a segregating population is obtained, whereas NILs can take years to generate (Michelmore *et al.*, 1991). Secondly, NILs often contain some residual chromosomal segments from the donor parent, whereas it is highly unlikely that bulked segregant analysis will identify unlinked markers if there is a sufficient number of

individuals in each bulk (Michelmore *et al.*, 1991). The bulked segregant analysis method was used to find markers associated with downy-mildew resistance genes in lettuce plants (Michelmore *et al.*, 1991). Bulked segregant analysis and the use of NILs allow for the rapid screening of a large number of markers, and result in the quick identification of markers tightly linked to chromosomal regions of interest, without requiring a genetic map that is highly saturated throughout the genome.

Even with the help of these two strategies however, the accuracy at which closely linked markers can be mapped relative to a gene of interest depends on the number of recombination events actually present in the mapping population. Churchill *et al.* (1993) describe the pooled-sampling method, which allows for the quick identification, in a large population, of individuals with recombination events adjacent to the target locus. If we consider once more the cross described above, a large number of the F_2 plants are scored for the trait of interest, and plants that are either homozygous dominant or recessive are grouped into several pools. DNA is extracted from each pool and subjected to analysis with markers that are known to be linked to the target locus. Based on the proportion of pools that show recombination between each marker and the locus of interest, high-resolution mapping can be achieved using conventional three-point analysis and the markers can be ordered relative to each other and to the gene of interest. The probability of ordering the markers correctly mainly depends on pool size, which should depend on marker density in the target region (Churchill *et al.*, 1993). In short, pooled sampling allows the incorporation of rare recombination events in high-resolution mapping studies, without having to extract DNA from and analyze every single F_2 plant in the population.

In contrast, if the objective is to roughly map loci throughout the genome to compare their map position with that of known loci, as is the case in this thesis, it is possible to use an informative subset of the mapping population (Mgonja *et al.*, 1994; Derek Lydiate, personal communication). A few individuals contribute most of the information in a mapping population. If a large number of markers are evenly distributed throughout the genome, sets of markers delineating key intervals can be selected and the minimum number of plants possessing recombination events in those intervals can be identified. This subset of plants can then be used to analyze the markers to be mapped. Once a marker maps to a particular interval, finer mapping can be accomplished by selecting more plants that show recombination events within that interval (Mgonja *et al.*, 1994).

1.2.2.2 Comparative Genome Mapping

As flowering plants have evolved from a common ancestor, they share conserved segments of DNA (Tanksley et al., 1995). Paterson *et al.* (1996) predict that 43-58% of DNA fragments less than 3 kb in size should remain co-linear in all flowering plants. They therefore suggest that it should eventually be possible to create a unified genetic map for higher plants. In fact, they estimate that *Arabidopsis* and *Sorghum* might differ by only 200 chromosomal rearrangements. On that note, they bring up the interesting point that, if similar genes confer resistance to the same pathogens in different crops, the practice of crop rotation, as a sustainable practice to avoid the development of resistance in the pathogen, may have to be re-considered.

The availability of a comparative map between two crops allows the pooling of the genetic data from both species, in the regions of colinearity (Tanksley et al., 1995). This should prove especially useful in the transfer of genetic information from well-characterized model plants, such as *Arabidopsis*, to less characterized crop plants such as the crop *Brassicas* (Lagercrantz *et al.*, 1996). This is especially true for the cloning of plant disease resistance genes. A number of R genes have been cloned in Arabidopsis (Bent *et al.*, 1994; Mindrinos *et al.*, 1994; Grant *et al.*, 1995; Parker *et al.*, 1997), and over 80 additional R genes and their homologues have been mapped (Wang *et al.*, 1996; Shauna Somerville, personal communication). This in sharp contrast to crops such as *Brassica napus* for which no resistance genes have been cloned, and only two have been mapped (Ferreira *et al.*, 1995a and b; Mayerhofer *et al.*, 1997; Dion *et al.*, 1995). The extensive genetic resources of *Arabidopsis*, and the comparative maps available for the *Brassicas* (Teutonico and Osborn, 1994; Lagercrantz *et al.*, 1996; Lagercrantz and Lydiate, 1996; Osborn *et al.*, 1997) may therefore be used to locate and eventually clone similar resistance genes in *Brassica napus* and other related crops.

1.2.3 Mapping of Plant Disease Resistance Gene Homologues

Given the extensive conservation of genes conferring resistance to very different pathogens in distantly related plants, and even in animals (see section 1.1.3.5 above), it should be able to identify additional resistance genes by looking for homology to known R genes.

Wang *et al.* (1996) first used this approach in *Arabidopsis*. They used a probe, corresponding to the predicted LRR domain of a cDNA clone with strong homology to

the *RPS2* and *RPM1* genes, to hybridize an *Arabidobsis* bacterial artificial chromosome (BAC) library. They found seven clones, corresponding to 4 different classes. This is a small number however, as it is estimated that *Arabidopsis* contains over 100 R genes (Shauna Somerville, personal communication). Presumably the use of hybridization, even at low stringencies, limits considerably the number of different R genes that can be isolated from a given probe.

Kanazin *et al.* (1996) and Yu *et al.*, (1996) looked for R gene homologues in soybean, but they used a PCR approach. They designed degenerate primers corresponding to conserved amino acid stretches in the nucleotide-binding sites of the *N*, *RPS2*, and L^6 genes, to amplify similar sequences in soybean. Kanazin *et al.* (1996) obtained a single PCR product, whereas Yu *et al.* (1996) obtained two, one of which was also present in rice, barley, tobacco, and potato. In both studies however, the PCR products investigated corresponded to different fragments of identical size. These products were grouped into 9 and 11 distinct classes (Kanazin *et al.*, 1996; and Yu *et al.*, 1996 respectively) that shared homology with each other and with known resistance genes. In both cases, some of these classes mapped to the vicinity of known disease resistance loci. Kanazin *et al.* (1996) found that their resistance gene analogs (RGA) were found either singly or in clusters throughout the genome. As mentioned earlier, this is in accordance with current data on plant R genes.

Though this PCR approach seems very promising for the identification of new R genes, both sets of authors warn against potential complicating factors. Kanazin *et al.* (1996) suggest that at least one of their RGA clones corresponds to a pseudogene, as it contains numerous stop codons. They mention that pseudogenes, because they are more

likely to accumulate mutations will be more easily mapped, and they suggest that mapping RGAs might be more useful in identifying R gene clusters for map-based cloning than in actually cloning new R genes directly. Yu *et al.* (1996) also point out that some genes with homology to R genes might not be involved in resistance at all. This seems to be the case at least for the *Arabidopsis* developmental genes *CLAVATA1* and *ERECTA*, which share homology with R gene serine/threonine kinases (see section 1.1.3.5 above).

In a similar study in potatoes, Leister D. *et al.* (1996) used three different primer pair combinations to amplify sequences homologous to the nucleotide-binding sites of the *N* and *RPS2* genes. Though the primers were very similar, they produced completely different products, suggesting that a large number of different R genes may be accessible by slightly modifying the primers used to isolate them. These primers produced DNA fragments that detected numerous loci that segregated with 14 known disease resistance loci in potato, tomato and tobacco. Another primer pair was designed to amplify *Pto* homologues. This primer pair detected a complex locus that mapped to a region homologous to the *Pto*-containing region of tomato. Finally, a *Cf-9* cDNA probe from tomato also detected a complex locus in potato, at a position homologous to that in tomato. Interestingly, the probes derived from this study detect similar fragments in a wide range of plants, in yeast, and even in humans (Leister D. *et al.*, 1996), confirming that homologous sequences are present in very divergent taxa.

Recently, Feuillet *et al.* (1997) isolated the first plant R gene to be cloned by homology. They used a wheat cDNA probe corresponding to a serine/threonine kinase to perform NIL analysis. They found a fragment that co-segregated with the rust resistance

gene *LR10*. Upon cloning of this fragment, they found that a gene, *LRK10*, which encodes a new type of plant receptor kinase. They did not perform complementation analysis however.

In conclusion, the use of sequence conservation in finding resistance gene homologues promises to yield numerous new candidate resistance genes, but there is still a long way before we actually obtain several cloned genes with defined resistance specificities (Michelmore, 1996). The ultimate proof that homologues correspond to functional R genes will come from complementation and mutant analysis (Michelmore, 1996).

Michelmore (1996) also warns against the exclusive use of the nucleotide-binding site in homology searches, as this domain is present in numerous different proteins. He suggests that the combination of a LRR and a nucleotide-binding site in the same protein is a better indicator of plant R genes. Using the LRR domain in PCR studies has been hindered however by the fact that its consensus sequence is very loose. This may change as new R genes are sequenced, and conserved patterns in LRR domains emerge (Michelmore, 1996). In any case, it is well recognized that resistance gene homologues will have immediate benefits, as they represent the ideal markers for disease resistance in marker-assisted selection schemes (Michelmore, 1996).

1.2.4 Random Genome Sequencing

Another means by which plant R genes can be isolated is through the random sequencing of plant genomes. As automated sequencing technologies are being developed, the speed at which genomes are being sequenced is increasing dramatically

(Gibson and Somerville, 1993). One of the applications of this technology is the sequencing of random cDNA ends, also called expressed sequence tags (ESTs) (Newman *et al.*, 1994). By comparing available *Arabidopsis* ESTs to sequences in the world DNA databases, over 80 plant R gene homologues were identified (Shauna Somerville, personal communication).

1.3 Summary

These are exciting times for the study of plant disease resistance. The increasing number of available plant disease resistance genes and pathogen avirulence genes, combined with the recent development of improved methods for gene mapping and cloning are allowing us to begin to dissect the gene-for-gene disease response hypothesized more than 50 years ago.

As genetic maps from genetically distant organisms are being integrated, we are starting to realize that some mechanisms, including disease resistance, might have been in place for over one billion years. Our increased knowledge of the mechanisms underlying disease resistance should allow us to engineer plants with novel combinations of resistance specificities, thus contributing to the global effort that is needed to secure food production worldwide, as we rapidly approach the third millenium and all the challenges it promises.

1.4 Objectives

The aim of this research was to isolate and map, in *Brassica napus* L., sequences with homology to known disease resistance genes. Specific objectives were:

- To amplify, clone, and sequence DNA regions from the canola genome with homology to known plant disease resistance genes.
- To map these sequences onto a *B. napus* genetic map (Ferreira *et al.*, 1994).
- To show that it is possible to map loci of interest using a small subset of a mapping population.
- To develop a PCR-based approach for quickly testing whether specific primer pairs, designed from R gene conserved domains, produce amplification products that map to a disease resistance locus of interest.

This project was intended to generate a database of putative disease resistance genes in *B. napus*. As breeders and geneticists identify more and more disease resistance loci, their genetic position will be compared to that of the sequences we have identified. Should any of these loci map to the same position as our sequences, the cloning of the corresponding resistance gene would be greatly facilitated.

Furthermore, the method for the quick mapping of resistance gene homologues that we have developed may facilitate the cloning of disease resistance genes for which the approximate chromosomal location is known, in *Brassica napus* or even other plant species. This is in an effort to improve the efficiency in isolating specific resistance genes based on homology with known R genes. In the long term, we hope that this project will contribute to our understanding of the disease response in plants, which would, in turn, increase our ability to create disease resistant crops more effectively.

1.5 Literature Cited

Agrios, G.N. (1988) Plant pathology. 3rd Edition. Academic Press. New York.

Akkaya, M.S., Bhagwat, A.A., and Cregan, P.B. (1992) Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics* **132**:1131-1139.

Alfano, J.R., and Collmer, A. (1997) The type III (Hrp) secretion pathway of plant pathogenic bacteria: trafficking harpins, Avr proteins, and death. *Journal of Bacteriology* **179**:5655-5662.

Allen, F.L. (1994) Usefulness of plant genome mapping to plant breeding. *In*: Plant Genome Analysis. pp. 11-18. Gresshoff, P.M. (Editor). CRC Press.

Anderson, P.A., Lawrence, G.J., Morrish, B.C., Ayliffe, M.A., Finnegan, E.J., and Ellis, J.G. (1997) Inactivation of the flax rust resistance gene *M* associated with loss of a repeated unit within the leucine-rich repeat coding region. *The Plant Cell* **9**:641-651.

Baker, B., Zambryski, P., Staskawicz, B., and Dinesh-Kumar, S.P. (1997) Signaling in plant-microbe interactions. *Science* **276**:726-733.

Barker, R.F., Thompson, D.V., Talbot, D.R., Swanson, J., and Bennetzen, J.L. (1984) Nucleotide sequence of the maize transposable element *Mu1*. *Nucleic Acids Research* **12**:5955-5967.

Belvin, M.P., and Anderson, K.V. (1996) A conserved signaling pathway: the *Drosophila* Toll-Dorsal pathway. *Annual Reviews of Cell and Developmental Biology* **12**:393-416.

Bent, A.F., Kunkel, B.N., Dahlbeck, D., Brown, K.L., Schmidt, R., Giraudat, J., Leung, J., and Staskawicz, B.J. (1994) *RPS2* of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. *Science* **265**:1856-1860.

Bi, Y., Kenton, P., Mur, L., Darby, R., and Draper, J. (1995) Hydrogen peroxide does not function downstream of salicylic acid in the induction of PR protein expression. *The Plant Journal* **8**:235-245.

Biffen, R.H. (1905) Mendel's laws of inheritance and wheat breeding. *Journal of Agricultural Science* **1**:4-48.

Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**:314-331.

Braun, T., Schofield, P.R., and Sprengel, R. (1991) Amino-terminal leucine-rich repeats in gonadotropin receptors determine hormone selectivity. *The EMBO Journal* **10**:1885-1890.

Cai, D., Kleine, M., Kifle, S., Harloff, H.J., Sandal, N.N., Marcker, K.A., Klein-Lankhorst, R.M., Salentijn, E.M.J., Lange, W., Stiekema, W.J., Wyss, U., Grundler, F.M.W., and Jung, C. (1997) Positional cloning of a gene for nematode resistance in sugar beet. *Science* **275**:832-834.

Cao, H., Glazebrook, J., Clarke, J.D., Volko, S., and Dong, X. (1997) The *Arabidopsis NPR1* gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats. *Cell* **88**:57-63.

Chandra, S., Martin, G.B., and Low, P.S. (1996) The Pto kinase mediates a signaling pathway leading to the oxidative burst in tomato. *Proceedings of the National Academy of Sciences USA* **93**:13393-13397.

Chen, Z., Silva, H., and Klessig, D.F. (1993) Active oxygen species in the induction of plant systemic acquired resistance by salicylic acid. *Science* **262**:1883-1886.

Chuck, G., Robbins, T., Nijjar, C., Ralston, E., Courtney-Gutterson, N., and Dooner, H.K. (1993) Tagging and cloning of a petunia flower color gene with the maize transposable element Activator. *The Plant Cell* **5**:371-378.

Churchill, G.A., Giovannoni, J.J., and Tanksley, S.D. (1993) Pooled-sampling makes high-resolution mapping practical with DNA markers. *Proceedings of the National Academy of Sciences USA* **90**:16-20.

Clark, S.E. (1997) The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in *Arabidopsis. Cell* **89**:575-585.

Dangl, J.L., Dietrich, R.A., and Richberg, M.H. (1996) Death don't have no mercy: cell death programs in plant-microbe interactions. *The Plant Cell* **8**:1793-1807.

Després, C., Subramaniam, R., Matton, D.P., and Brisson, N. (1995) The activation of the potato *PR-10a* gene requires the phosphorylation of the nuclear factor PBF-1. *The Plant Cell* **7**:589-598.

Dion, Y., Gugel, R.K., Rakow, G.F.W., Séguin-Swartz, G., and Landry, B.S. (1995) RFLP mapping of resistance to the blackleg disease [causal agent, *Leptosphaeria maculans* (Desm.) Ces. et de Not.] in canola (*Brassica napus* L.). *Theoretical and Applied Genetics* **91**:1190-1194.

Dixon, M.S., Jones, D.A., Keddie, J.S., Thomas, C.M., Harrison, K., and Jones, J.D.G. (1996) The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell* **84**:451-459.

Durner, J., and Klessig, D.F. (1995) Inhibition of ascorbate peroxidase by salicylic acid and 2,6-dichloroisonicotinic acid, two inducers of plant defense responses. *Proceedings of the National Academy of Sciences USA* **92**:11312-11316.

Durner, J., and Klessig, D.F. (1996) Salicylic acid is a modulator of tobacco and mammalian catalases. *Journal of Biological Chemistry* **271**:28492-28501.

Ellis, J.G., Lawrence, G.J., Finnegan, E.J., and Anderson, P.A. (1995) Contrasting complexity of two rust resistance loci in flax. *Proceedings of the National Academy of Sciences USA* **92**:4185-4188.

Fedoroff, N., Wessler, S., and Shure, M. (1983) Isolation of the transposable maize controlling elements *Ac* and *Ds*. *Cell* **35**:235-242.

Fedoroff, N., Furtek, D.B., and Nelson, O.E. Jr. (1984) Cloning of the bronze locus in maize by a simple and generalizable procedure using the transposable controlling element *Activator* (*Ac*). *Proceedings of the National Academy of Sciences USA* **81**:3825-3829.

Ferreira, M.E., Williams, P.H., and Osborn, T.C. (1994) RFLP mapping of *Brassica napus* using doubled haploid lines. *Theoretical and Applied Genetics* **89**:615-621.

Ferreira, M.E., Rimmer, S.R., Williams, P.H., and Osborn, T.C. (1995a) Mapping loci controlling *Brassica napus* resistance to *Leptosphaeria maculans* under different screening conditions. *Phytopathology* **85**:213-217.

Ferreira, M.E., Williams, P.H., and Osborn, T.C. (1995b) Mapping a locus controlling resistance to *Albugo candida* in *Brassica napus* using molecular markers. *Phytopathology* **85**:218-220.

Feuillet, C., Schachermayr, G., and Keller, B. (1997) Molecular cloning of a new receptor-like kinase gene encoded at the *Lr10* disease resistance locus of wheat. *The Plant Journal* **11**:45-52.

Flor, H.H. (1946) Genetics of pathogenicity in *Melamspora lini*. *Journal of Agricultural Research* **73**:335-357.

Flor, H.H. (1971) Current status of the gene-for-gene concept. *Annual Reviews of Phytopathology* **9**:275-296.

Gibson, S., and Somerville, C. (1993) Isolating plant genes. *Trends in Biotechnology* **11**:306-313.

Giovannoni, J.J., Wing, R.A., Ganal, M.W., and Tanksley, S.D. (1991) Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucleic Acids Research* **19**:6553-6558.

Gopalan, S., Bauer, D.W., Alfano, J.R., Loniello, A.O., He, S.Y., and Collmer, A. (1996) Expression of *the Pseudomonas syringae* avirulence protein AvrB in plant cells alleviates its dependence on the hypersensitive response and pathogenicity (Hrp) secretion system in eliciting genotype-specific hypersensitive cell death. *The Plant Cell* **8**:1095-1105.

Grant, M.R., Godiard, L., Straube, E., Ashfield, T., Lewald, J., Sattler, A., Innes, R.W., and Dangl, J.L. (1995) Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. *Science* **269**:843-846.

Greenberg, J.T., Guo, A., Klessig, D.F., and Ausubel, F.M. (1994) Programmed cell death in plants: a pathogen-triggered response activated coordinately with multiple defense functions. *Cell* **77**:551-563.

Hammond-Kosack, K.E., Harrison, K., and Jones, J.D.G. (1994) Developmentally regulated cell death on expression of the fungal avirulence gene *Avr9* in tomato seedlings carrying the disease resistance gene *Cf-9*. *Proceedings of the National Academy of Sciences USA* **91**:10445-10449.

Jackson, A.O., and Taylor, C.B. (1996) Plant-microbe interactions: life and death at the interface. *The Plant Cell* **8**:1651-1668.

Jia, Y., Loh, Y.T., Zhou, J., and Martin, J.B. (1997) Alleles of *Pto* and *Fen* occur in bacterial speck-susceptible and fenthion-insensitive tomato cultivars and encode active protein kinases. *The Plant Cell* **9**:61-73.

Johal, G.S., and Briggs, S.P. (1992) Reductase activity encoded by the *HM1* disease resistance gene in maize. *Science* **258**:985-987.

Jones, D.A., Thomas, C.M., Hammond-Kosack, K.E., Balint-Kurti, P.J., and Jones, J.D.G. (1994) Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. *Science* **266**:789-793.

Kanazin, V., Marek, L.M., and Shoemaker, R.C. (1996) Resistance gene analogs are conserved and clustered in soybean. *Proceedings of the National Academy of Sciences USA* **93**:11746-11750.

Keen, N.T. (1990) Gene-for-gene complementarity in plant-pathogen interactions. *Annual Reviews of Genetics* **24**:447-463.

Keen, N.T. (1992) The molecular biology of disease resistance. *Plant Molecular Biology* **19**:109-122.

Kobayashi, D.Y., Tamaki, S.J., and Keen, N.T. (1989) Cloned avirulence genes from the tomato pathogen Pseudomonas syringae pv. tomato confer cultivar specificity on soybean. *Proceedings of the National Academy of Sciences USA* **86**:157-161.

Kobe, B., and Deisenhofer, J. (1993) Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature* **366**:751-756.

Kopp, E., and Ghosh, S. (1994) Inhibition of NF-κB by sodium salicylate and aspirin. *Science* **265**:956-958.

Kvaratskhelia, M., George, S.J., and Thorneley, R.N.F. (1997) Salicylic acid is a reducing substrate and not an effective inhibitor of ascorbate peroxidase. *Journal of Biological Chemistry* **272**:20998-21001.

Lagercrantz, U., Putterill, J., Coupland, G., and Lydiate, D.J. (1996) Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. *The Plant Journal* **9**:13-20.

Lagercrantz, U., and Lydiate, D.J. (1996) Comparative genome mapping in *Brassica*. *Genetics* **144**:1903-1910.

Lawrence, G.J., Finnegan, E.J., Ayliffe, M.A., and Ellis, J.G. (1995) The *L6* gene for flax rust resistance is related to the *Arabidopsis* bacterial resistance gene *RPS2* and the tobacco viral resistance gene *N*. *The Plant Cell* **7**:1195-1206.

Leister, D., Ballvora, A., Salamini, F., and Gebhardt, C. (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. *Nature Genetics* **14**:421-428.

Leister, R.T., Ausubel, F.M., and Katagiri, F. (1996) Molecular recognition of pathogen attack occurs inside of plant cells in plant disease resistance specified by the Arabidopsis genes *RPS2* and *RPM1*. *Proceedings of the National Academy of Sciences USA* **93**:15497-15502.

Levine, A., Tenhaken, R., Dixon, R., and Lamb, C. (1994) H_2O_2 from the oxidative burst orchestrates the plant hypersensitive disease resistance response. *Cell* **79**:583-593.

Loh, Y., and Martin, G.B. (1995) The disease-resistance gene *Pto* and the fenthion-sensitivity gene *Fen* encode closely related functional protein kinases. *Proceedings of the National Academy of Sciences USA* **92**:4181-4184.

Mann, C. (1997) Reseeding the green revolution. *Science* **277**:1038-1043.

Martin, G.B., Williams, J.G., and Tanksley, S.D. (1991) Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. *Proceedings of the National Academy of Sciences USA* **88**:2336-2340.

Martin, G.B., Brommonschenkel, S.H., Chuwongse, J., Frary, A., Ganal, M.W., Spivey, R., Wu, T., Earle, E.D., and Tanksley, S.D. (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* **262**:1432-1436.

Mayerhofer, R., Bansal, V.K., Thiagarajah, M.R., Stringam, G.R., and Good, A.G. (1997) Molecular mapping of resistance to *Leptosphaeria maculans* in Australian cultivars of *Brassica napus*. *Genome* **40**:294-301.

Meyer, M., Schreck, R., and Baeuerle, P.A. (1993) H_2O_2 and antioxidants have opposite effects on activation of NF-kappaB and AP-1 in intact cells: AP-1 as secondary antioxidant-responsive factor. *The EMBO Journal* **12**:2005-2015.

Mgonja, M.A., Dahleen, L.S., and Franckowiak, J.D. (1994) Subsets from mapping populations for localization of new genes in barley. *Barley Genetics Newsletter* **24**:14-23.

Michelmore, R.W., Paran, I., and Kesseli, R.V. (1991) Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences USA* **88**:9828-9832.

Michelmore, R. (1996) Flood warning - resistance genes unleashed. *Nature Genetics* **14**:376-378.

Mindrinos, M., Katagiri, F., Yu, G., and Ausubel, F.M. (1994) The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell* **78**:1089-1099.

Moffat, A.S. (1994) Mapping the sequence of disease resistance. *Science* **265**:1804-1805.

Newman, T., de Bruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel, E., and Somerville, C. (1994) Genes galore: a summary of methods for accessing results

from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Physiology* **106**:1241-1255.

Ochman, H., Gerber, A.S., and Hartl, D.L. (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**:621-623.

Ori, N., Eshed, Y., Paran, I., Presting, G., Aviv, D., Tanksley, S., Zamir, D., and Fluhr, R. (1997) The *I2C* family from the wilt disease resistance locus *I2* belongs to the nucleotide binding, leucine-rich repeat superfamily of plant resistance genes. *The Plant Cell* **9**:521-532.

Osborn, T.C., Kole, C., Parkin, I.A.P., Sharpe, A.G, Kuiper, M., Lydiate, D.J., and Trick, M. (1997) Comparison of flowering time genes in *Brassica rapa*, *B. napus*, and *Arabidopsis thaliana*. *Genetics* **146**:1123-1129.

Osbourn, A.E. (1996) Preformed antimicrobial compounds and plant defense against fungal attack. *The Plant Cell* **8**:1821-1831.

Parker, J.E., and Coleman, M.J. (1997) Molecular intimacy between proteins specifying plant-pathogen recognition. *Trends in the Biochemical Sciences* **22**:291-296.

Parker, J.E., Coleman, M.J., Szabò, V., Frost, L.N., Schmidt, R., Van der Biezen, E.A., Moores, T., Dean, C., Daniels, M.J., and Jones, J.D.G. (1997) The *Arabidopsis* downy mildew resistance gene *RPP5* shares similarity to the Toll and Interleukin-1 receptors with *N* and *L6*. *The Plant Cell* **9**:879-894.

Paterson, A.H., Lan, T., Reischmann, K.P., Chang, C., Lin., Y., Liu, S., Burow, M.D., Kowalski, S.P., Katsar, C.S., DelMonte, T.A., Feldmann, K.A., Schertz, K.F., and Wendel, J.F. (1996) Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence. *Nature Genetics* **14**:380-382.

Person, C., Samborski, D.J., and Rohringer, R. (1962) The gene-for-gene concept. *Nature* **194**:561-562.

Pring, D.R., and Lonsdale, D.M. (1989) Cytoplasmic male sterility and maternal inheritance of disease susceptibility in maize. *Annual Reviews of Phytopathology* **27**:483-502.

Rafalski, J.A., and Tingey, S.V. (1993) Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. *Trends in Genetics* **9**:275-280.

Rahme, L.G., Stevens, E.J., Wolfort, S.F., Shao, J., Tompkins, R.G., and Ausubel, F.M. (1995) Common virulence factors for bacterial pathogenicity in plants and animals. *Science* **268**:1899-1902.

Rao, G.N., Lasségue, B., Griendling, K.K., Alexander, R.W., and Berk, B.C. (1993) Hydrogen peroxide-induced c-fos expression is mediated by arachidonic acid release: role of protein kinase C. *Nucleic Acids Research* **21**:1259-1263.

Richter, T.E., Pryor, T.J., Bennetzen, J.L., and Hulbert, S.H. (1995) New rust resistance specificities associated with recombination in the *Rp1* complex in maize. *Genetics* **141**:373-381.

Rommens, C.M.T., Salmeron, J.M., Oldroyd, G.E.D., and Staskawicz, B.J. (1995) Intergeneric transfer and functional expression of the tomato disease resistance gene *Pto*. *The Plant Cell* **7**:1537-1544.

Rosqvist, R., Hakansson, S., Forsberg, A., and Wolf-Watz, H. (1995) Functional conservation of the secretion and translocation machinery for virulence proteins of yersiniae, salmonellae and shigellae. *The EMBO Journal* **14**:4187-4195.

Ryals, J.A., Neuenschwander, U.H., Willits, M.G., Molina, A., Steiner, H., and Hunt, M.D. (1996) Systemic acquired resistance. *The Plant Cell* **8**:1809-1819.

Ryals, J.A., Weymann, K., Lawton, K, Friedrich, L., Ellis, D., Steiner, H., Johnson, J., Delaney, T.P., Jesse, T., Vos, P., and Uknes, S. (1997) The Arabidopsis *NIM1* protein shows homology to the mammalian transcription factor inhibitor IκB. *The Plant Cell* **9**:425-439.

Salmeron, J.M., Oldroyd, G.E.D., Rommens, C.M.T., Scofield, S.R., Kim, H., Lavelle, D.T., Dahlbeck, D., and Staskawicz, B.J. (1996) Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. *Cell* **86**:123-133.

Schreck, R., Rieber, P., and Baeuerle, P.A. (1991) Reactive oxygen intermediates as apparently widely used messengers in the activation of the NF-kappaB transcription factor and HIV-1. *The EMBO Journal* **10**:2247-2258.

Scofield, S.R., Tobias, C.M., Rathjen, J.P., Chang, J.H., Lavelle, D.T., Michelmore, R.W., and Staskawicz, B.J. (1996) Molecular basis of gene-for-gene specificity in bacterial speck disease of tomato. *Science* **274**:2063-2065.

Shah, J., and Klessig, D.F. (1996) Identification of a salicylic acid-responsive element in the promoter of the tobacco pathogenesis-related β-1,3-glucanase gene, *PR-2d*. *The Plant Journal* **10**:1089-1101.

Song, W., Wang, G., Chen, L., Kim, H., Pi, L., Holsten, T., Gardner, J., Wang, B., Zhai, W., Zhu, L., Fauquet, C., and Ronald, P. (1995) A receptor kinase-like protein encoded by the rice disease resistance gene *Xa21*. *Science* **270**:1804-1806.

Stakman, E.C. (1914) A study in cereal rusts: Physiological races. *University of Minnesota, Agricultural Experiment Station Bulletin* **138**.

Staskawicz, B.J., Ausubel, F.M., Baker, B.J., Ellis, J.G., and Jones, J.D.G. (1995) Molecular genetics of plant disease resistance. *Science* **268**:661-667.

Staskawicz, B.J., Dahlbeck, D., and Keen, N.T. (1984) Cloned avirulence gene of *Pseudomonas syringae* pv. *glycinea* determines race-specific incompatibility on *Glycine max* (L.) Merr. *Proceedings of the National Academy of Sciences USA* **81**:6024-6028.

Subramaniam, R., Després, C., and Brisson, N. (1997) A functional homolog of mammalian protein kinase C participates in the elicitor-induced defense response in potato. *The Plant Cell* **9**:653-664.

Sudupak, M.A., Bennetzen, J.L., and Hulbert, S.H. (1993) Unequal exchange and meiotic instability of disease-resistance genes in the *Rp1* region of maize. *Genetics* **133**:119-125.

Tang, X., Frederick, R.D., Zhou, J., Halterman, D.A., Jia, Y., and Martin, G.B. (1996) Initiation of plant disease resistance by physical interaction of AvrPto and Pto kinase. *Science* **274**:2060-2062.

Tanksley, S.D., Ganal, M.W., and Martin, G.B. (1995) Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. *Trends in Genetics* **11**:63-68.

Tanksley, S.D., and McCouch, S.R. (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* **277**:1063-1066.

Teutonico, R.A., and Osborn, T.C. (1994) Mapping of RFLP and qualitative trait loci in *Brassica rapa* and comparison to the linkage maps of *B. napus*, *B. oleracea*, and *Arabidopsis thaliana*. *Theoretical and Applied Genetics* **89**:885-894.

Thilmony, R.L., Chen, Z., Bressan, R.A., and Martin, G.B. (1995) Expression of the tomato *Pto* gene in tobacco enhances resistance to *Pseudomonas syringae* pv *tabaci* expressing *avrPto*. *The Plant Cell* **7**:1529-1536.

Torii, K.U., Mitsukawa, N., Oosumi, T., Matsuura, Y., Yokoyama, R., Whittier, R.F., and Komeda, Y. (1996) The *Arabidopsis Erecta* gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. *The Plant Cell* **8**: 735-746.

Van den Ackerveken, G., Marois, E., and Bonas, U. (1996) Recognition of the bacterial avirulence protein AvrBs3 occurs inside the host plant cell. *Cell* **87**:1307-1316.

Vanderplank, J.E. (1963) Plant diseases: Epidemics and control. Academic Press. New York.

Wang, G., Warren, R., Innes, R, Osborne, B., Baker, B., and Ronald, P.C. (1996) Construction of an *Arabidopsis* BAC library and isolation of clones hybridizing with disease-resistance, gene-like sequences. *Plant Molecular Biology Reporter* **14**:107-114.

White, R.F. (1979) Acetylsalicylic acid (aspirin) induces resistance to tobacco mosaic virus in tobacco. *Virology* **99**:410-412.

Whitham, S., Dinesh-Kumar, S.P., Choi, D., Hehl, R., Corr, C., and Baker, B. (1994) The product of tobacco mosaic virus resistance gene *N*: similarity to Toll and the Interleukin-1 receptor. *Cell* **78**:1101-1115.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., and Tingey, S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**:6531-6535.

Williams, J.G.K., Reiter, R.S., Young, R.M., and Scolnik, P.A. (1993) Genetic mapping of mutations using phenotypic pools and mapped RAPD markers. *Nucleic Acids Research* **21**:2697-2702.

Yang, Y., and Gabriel, D.W. (1995) Intragenic recombination of a single plant pathogen gene provides a mechanism for the evolution of new host specificities. *Journal of Bacteriology* **177**:4963-4968.

Yu, Y.G., Buss, G.R., and Saghai-Maroof, M.A. (1996) Isolation of a superfamily of candidate disease resistance genes in soybean based on a conserved nucleotide-binding site. *Proceedings of the National Academy of Sciences USA* **93**:11751-11756.

Zhou, J., Loh, Y.T., Bressan, R.A., and Martin, G.B. (1995) The tomato gene *Pti1* encodes a serine/threonine kinase that is phosphorylated by *Pto* and is involved in the hypersensitive response. *Cell* **83**:925-935.

Zhou, J., Tang, X., and Martin, G.B. (1997) The Pto kinase conferring resistance to tomato bacterial speck disease interacts with proteins that bind a cis-element of pathogenesis-related genes. *The EMBO Journal* **16**:3207-3218.

2.0 Genetic Mapping of Plant Resistance Gene Homologues in *Brassica napus* L., using a minimal mapping population

2.1 Introduction

Canola cultivars of *B. napus* are of high economical importance to Canada, and breeding for this crop has been intensive since the first cultivar was licensed in 1974 (Adolphe, 1980). However, canola is still highly susceptible to a number of diseases and insect pests. The blackleg fungus (*Leptosphaeria maculans*) alone has caused over 50 million dollars of damage annually for the 1985-1995 period, across the Canadian Prairies (Kharbanda *et al.*, 1995). This problem is compounded by the fact that, like most cash crops, canola is grown in large monocultures. Pathogens are therefore under high selection pressure to evolve new virulence genes to counteract any defense mechanisms that may have been bred into the canola cultivars in use.

There is increasing interest in the "pyramiding" of disease resistance genes to reduce the pest problems associated with monoculture. This approach however is very labor and time intensive when traditional selection techniques are used. Gene pyramiding is greatly simplified when closely linked molecular markers are available for the desired resistance genes. The best markers for disease resistance, obviously, would be the resistance genes themselves (Michelmore, 1996). To this date, the most limiting factor in the use of marker-assisted selection for disease resistance gene pyramiding in *Brassica napus* is the number of mapped resistance loci. To the best of our knowledge, only three sources of resistance to the blackleg fungus (Dion *et al.*, 1995; Ferreira *et al.*, 1995a; Mayerhofer *et al.*, 1997) and one source of resistance to white rust (Ferreira *et al.*, 1995b) have been mapped in *B. napus*, and the corresponding genes have not yet been cloned.

Over the last five years, a plethora of genes conferring resistance to a large variety of diseases in many different crops have been cloned (reviewed in Baker *et al.*, 1997). As they were being cloned, it became evident that there were a number of conserved motifs that characterized these resistance genes. In particular, most of them possess leucine-rich repeats (LRR) that are thought to be involved in protein-protein interactions (Baker *et al.*, 1997). One class of resistance genes, including the *Arabidopsis thaliana RPS2* gene, possesses both an LRR and a nucleotide-binding site (NBS) (Bent *et al.*, 1994; Mindrinos *et al.*, 1994). We used this sequence conservation to amplify and map similar sequences in *B. napus* which, like *Arabidopsis*, is a member of the Brassicaceae. The sequences we have mapped constitute a database of putative disease resistance genes in *B. napus*, which may segregate with disease resistance loci that will be mapped in the future, thus providing physical anchors to the corresponding resistance genes and/or markers for use in marker-assisted selection.

Several groups have already used this approach in soybean and potatoes, and have found sequences that segregate with known disease resistance loci (Kanazin *et al.*, 1996; Yu *et al.*, 1996; Leister *et al.*, 1996). However, to date, nobody has utilized both the leucine-rich repeats and the NBS domains in their search for R gene homologues. Michelmore (1996) pointed out that there are many types of NBS-containing proteins, and that the presence of both NBS and LRR domains in the same protein is a more stringent criterion for the identification of R gene homologues.

Not all resistance genes are of the NBS/LRR class however, and many economically important resistance genes will presumably escape our attention if we concentrate all our efforts on this class. Also, though we are likely to be flooded with R

gene homologues in the near future, it will be some time before we find out which ones correspond to functional R genes, as the molecular data is quickly surpassing the genetic data (Michelmore, 1996). Thus there is a need for a method to rapidly test putative R genes to discern which ones will have an immediate use.

We have designed a method to map disease resistance gene homologues quickly, using the polymerase chain reaction (PCR). It will therefore be possible to design primers corresponding to conserved domains of different classes of resistance genes and quickly test whether they amplify sequences that co-segregate with mapped resistance loci. In addition, Leister *et al.* (1996) found that PCR primers differing by as few as 2 nucleotides amplify very different R gene homologues, which suggests that various different primers should be designed for every conserved domain, so as to find the maximum number of interesting sequences. Once again, the method described in this thesis should allow the quick testing of these primers.

As putative disease resistance loci are mapped throughout the genomes of many important crop plants, and as our knowledge of the structure/function relationship in disease resistance genes develops, it will become easier to produce crops with the appropriate combination of disease resistance specificities.

Traditionally, plant genetic mapping has relied on the analysis of large mapping populations. However, in such populations, a few individuals contribute most of the information (Mgonja *et al.*, 1994). If there is a large number of markers on the genetic map, providing a good coverage of the genome, it is possible to identify a small number of individuals from the mapping population that possess recombination events within critical intervals (Mgonja *et al.*, 1994). From then on, it should be possible to assign new

markers to these intervals using only the chosen subset of the mapping population (Mgonja *et al.*, 1994). Though many researchers use subsets of their mapping populations for genetic mapping, the validity of this practice, to our knowledge, has not yet been formally demonstrated. We defined a minimal *Brassica napus* mapping population of 23 individuals and set out to validate this method.

2.2 Materials and Methods

2.2.1 Plant materials

Seeds from *B. napus* cultivars Major and Stellar, and 105 doubled-haploid lines derived from the F_1 of this cross, were generously provided by Dr. Tom Osborn (Ferreira *et al.*, 1994). Major is a biennial cultivar, requiring a vernalization period, while Stellar is an annual cultivar. We chose this mapping population, as it was used to map both disease resistance loci identified in *B. napus* so far (Ferreira *et al.*, 1995a and b). All plants were grown in a growth chamber with a 16 hour photoperiod, 65% relative humidity, a 20°C day temperature, and a 15°C night temperature. For seed increases of the Major x Stellar population, the plants were germinated in the growth chamber for approximately one week, transferred to a cold room at 4°C for 10 weeks, and finally transferred back to the growth chamber. Leaves of each plant were removed and either frozen at –80°C, lyophilized at –40 °C and stored at room temperature, or they were immediately ground in liquid nitrogen and used right away for DNA extraction.

2.2.2 Genomic DNA Extractions

For PCR, plant genomic DNA was prepared from 0.3g of lyophilized tissue using the CTAB method described in Doyle and Doyle (1987), except that the extraction buffer

contained 3% (w/v) CTAB, and 0.4% (v/v) β-mercaptoethanol. Alternatively, for quick mapping using PCR, genomic DNA was extracted from fresh leaf discs using a modification of the protocol described in Edwards *et al.* (1991). Leaf discs were macerated at room temperature with sterile pestles. Four hundred μL of extraction buffer [200mM Tris pH 7.5, 250mM NaCl, 25mM EDTA, and 0.5% SDS (w/v)] was added to each tube, and the solutions were centrifuged for 1 minute at full speed in a microcentrifuge. 300μL of supernatant were then added to an equal volume of isopropanol. The tubes were vortexed and left at room temperature for 2 minutes. This was followed by a 5 minute centrifugation. Finally, the pellets were washed with 500μL 70% ethanol (v/v), air-dried, and resuspended in 100μL TE buffer at 4°C overnight. Five μL of each preparation were used in each PCR reaction. For Southern hybridizations, genomic DNA was extracted using the Kirby method described in Sharpe et al. (1995), except that 5-7g of fresh tissue was ground in liquid nitrogen. All DNA was quantified using the Hoechst 33258 dye method and a Sequoia 400 fluorometer, as described in Ausubel *et al.* (1989).

2.2.3 Amplification and subcloning of disease resistance gene homologues

Primers were designed using the computer program Oligo v.4.1 for IBM (see Table 2.1). The forward primer was designed to correspond to the P-loop of the nucleotide-binding site (NBS) of the *A. thaliana RPS2* gene (Bent *et al.*, 1994; Mindrinos *et al.*, 1994), starting at nucleotide position 552. Two reverse primers were designed, corresponding to two regions of the *RPS2* leucine-rich repeat motif. The 3' end of reverse primer number one is designed to bind the RPS2 gene at nucleotide position 1849, while

the 3' end of primer number two binds at position 1588. The two primer pair combinations were used to amplify *B. napus* cv. Quantum genomic DNA, under three different sets of reaction conditions. PCR reactions were setup in a 25µL volume, with 30ng of genomic DNA, 200µM each of dNTPs, 50mM KCl, 10mM Tris-Cl pH 9.0, 0.01% (w/v) gelatin, 0.1% (v/v) Triton X-100, and 2.5U Taq polymerase. Primer and $MgCl_2$ concentrations varied, depending on the experiment (see Table 2.2).

Table 2.1. Primers for amplification of resistance gene homologues in the Brassicaceae

Primer Sequence	T_m[a]
Forward : 5' –TGGGGTGGGAAGACAAC	50.0
Reverse 1[b]: 5' – CTCGTGGGATCGTCTGAAG	50.1
Reverse 2[c]: 5' – GTCAGTTTCGGGCATATGAG	49.6

[a] Annealing temperature, under conditions of 250 pmol of oligonucleotide, and 50 mM KCl.
[b] The forward and reverse 1 primers are designated as the PLRR-1 primer pair (for P-loop/LRR).
[c] The forward and reverse 2 primers are designated as the PLRR-2 primer pair.

Table 2.2. PCR Reaction Conditions for the Amplification of R-gene Homologues.

	Experiment A	*Experiment B*	*Experiment C*
$MgCl_2$ Concentration	1.5 mM	4.5 mM	1.5 mM
Primer Concentration	0.125 µM (each)	0.5 µM	0.125 µM
Annealing Temperature	54°C	54°C	50°C

The reactions were performed in a PerkinElmer 9600 thermal cycler. A 3-minute period at 94°C was followed by 35 cycles of 1 minute at 94°C, 1 minute at 54°C, and 3 minutes at 72°C. This was followed by a final 5-minute elongation period at 72°C. The tubes were then held at 4°C or frozen at –20°C until they were used. The annealing temperature was also varied, depending on the experiment (see Table 2.2).

PCR products from all three sets of amplification conditions were blunt-ended and purified as described in Wang *et al.* (1994). The products were then ligated into the *Eco*RV site of pBluescript II KS+ (Stratagene Inc.) and the resulting plasmids were used

to transform *Escherichia coli* DH5α cells. The clones were grown on LB agar plates with 100μg/mL ampicillin, 40 mg/mL Xgal and 40 mg/mL IPTG. Plasmid DNA was extracted from positive clones using a standard alkaline lysis method, digested with the restriction enzyme *Rsa*1 (Pharmacia), and separated in a 1% (w/v) agarose gel. Based on the restriction patterns obtained, the clones were grouped into 27 families. All microbiological techniques and plasmid DNA manipulations were performed as described in Sambrook *et al.* (1989).

2.2.4 DNA Sequencing

All sequencing reactions were performed using an ABI Prism 377 automated DNA sequencer, according to manufacturer's instructions. Plasmid DNA was prepared using either a Pharmacia GFX™ Micro Plasmid Prep kit or according to the ABI Prism DyeDeoxy Termination Cycle Sequencing protocol. The inserts were sequenced using the T3 and T7 primers of the pBluescript KS+ plasmid.

2.2.5 Southern hybridizations

Plant genomic DNA from both parental lines and all 23 DH lines of the minimal mapping population was digested with the enzymes *Eco*RI, *Eco*RV, *Bam*HI, *Hind*III, *Hae*III, *Rsa*I, and *Xba*I (Pharmacia). All digestions were carried out in One-Phor-All+ buffer (Pharmacia) with 5μg DNA, 50U enzyme, and 2.67μM spermidine, at 37°C for 4 hours. The DNA fragments were separated overnight in a 0.8% agarose gel and transferred to GeneScreen*Plus* hybridization membranes (DuPont) using a vacuublotter (BRL), according to manufacturer's instructions. Membranes were baked at 80°C for 2 hours, and stored at room temperature until they were used in hybridizations.

The insert from one clone of each of the 27 families defined above was labelled with [^{32}P]dCTP, as described in Sharpe *et al.* (1995). Inserts were amplified using the T3 and T7 primers. PCR reactions were carried out as described above in section 2.2.3, except that the annealing temperature was 65°C, and the elongation step was only 1 minute. Two µL of each PCR reaction were used in the labeling reaction. After labeling, all probes were purified using Sephadex G-50 Nick Columns (Pharmacia).

Hybridizations were performed as described in Sharpe et al. (1995), except that 7.5% dextran sulfate, and 0.01% herring sperm DNA were used. The membranes were washed twice in 2xSSC; 0.1% SDS (w/v) at 65°C.

2.2.6 Quick genetic mapping using the polymerase chain reaction

DNA from both parental lines was amplified using both primer pairs and PCR reaction conditions "A" (stringent) and "B" (relaxed) described above. 15 µL of amplification products were digested in One-Phor-All+ buffer (Pharmacia), with 5U of one of the following enzymes: *Eco*RI, *Eco*RV, *Sau*3AI, *Pst*I, *Hha*I, *Dde*I, *Hae*III, and *Hpa*II. Polymorphisms between Major and Stellar where further investigated by amplifying DNA from the minimal mapping population with the corresponding primer pair, and digesting it with the appropriate enzyme.

2.2.7 Linkage analysis

Mapping of the resistance gene homologues was performed with the help of MAPMAKER/EXP 3.0b (Lander *et al.* 1987). All 480 markers were used in the analysis, but only the 23 lines of the minimal population were included at first. The command "Try" was used to assign a position to our markers within the defined sequence of

markers. When more than one position was assigned, each position was tested by gradually increasing the LOD score threshold. In some cases, it was not possible to distinguish between alternative map positions. In those cases, all map positions are reported and the results are discussed in section 2.4.

Once a map position was assigned to a particular sequence, the genetic distances between it and the surrounding markers were estimated using the information from all 105 DH lines of the mapping population.

2.2.8 Selection and validation of the minimal mapping population

A mapping data set showing the scoring for 480 markers in the Major x Stellar population was generously provided by Dr. Tom Osborn. From that data set, markers were selected every 10-15 cM. At times however, the gap between two markers was larger than 15 cM. The new data set was re-printed, and all the recombination events in the population were identified as a change in allele from one marker to the next. The doubled haploid line with the largest number of recombination events was chosen. Lines were then added to the minimal mapping population, as to represent one recombination event between each of the selected markers. When all the recombination events were represented, lines in the minimal population that were found to provide only redundant information were eliminated. This left us with 23 DH lines that, together, represented at least one recombination event every 10-15 cM. These lines are: 302, 311, 312, 314, 319, 321, 331, 338, 340, 343, 346, 364, 365, 367, 370, 386, 389, 395, 398, 402, 406, 407, 425. We then created a data set with all 480 markers for the 23 chosen lines. We analyzed this data set with MAPMAKER, using the command "LOD Table". The output generated

consists of a table of LOD scores and genetic distances for each marker pair combination in the sequence specified. For example, a marker compared to itself should give a distance of 0.0 cM and a high LOD score (typically 10-14). We performed this analysis on 5 random sets of 10 consecutive markers in the analysis.

2.3 Results

2.3.1 Resistance gene homologues in the Brassicaceae

Amplification of *B. napus* and *A. thaliana* genomic DNA using our PLRR primers under stringent and relaxed conditions is shown in Figure 2.1. Under stringent conditions, both PLRR-1 and 2 primer sets give a strong main band. The expected size of these bands, based on the *A. thaliana RPS2* sequence, is 1316 bp and 1056 bp for PLRR-1 and PLRR-2 respectively. Though the main band obtained for *B. napus*, with PLRR-2, approaches this expected size very closely, the band obtained with PLRR-1 is much smaller than expected. Relaxed conditions produce, as expected, several additional bands, both smaller and larger than the main band. We later found that inserts of the same size could actually belong to different classes, based on their *Rsa*I restriction patterns. In all cases, some bands were found to be unique to one of the two species.

Amplification products are seen in all species of the Brassicaceae examined, when our PCR primers PLRR-2 are used (Figure 2.2). The main band found in *B. napus* is also found in 6 of the 7 other species examined, including the more distantly related *Sinapis arvensis*. However, the banding patterns are most similar between *B. napus* and *B. rapa*. *B. oleracea* possesses an additional lower band. *B. nigra*, which is thought to have hybridized with *B. oleracea* and *B. rapa* to form *B. carinata* and *B. juncea* respectively

(U, 1935), shares a band, just under the 0.7 kb mark, with the latter two. *B. nigra* also shows a unique band at approximately 1.9 kb. Amplification of *S. alba* DNA with PLRR-2 produced only one band, also present in *S. arvensis*.

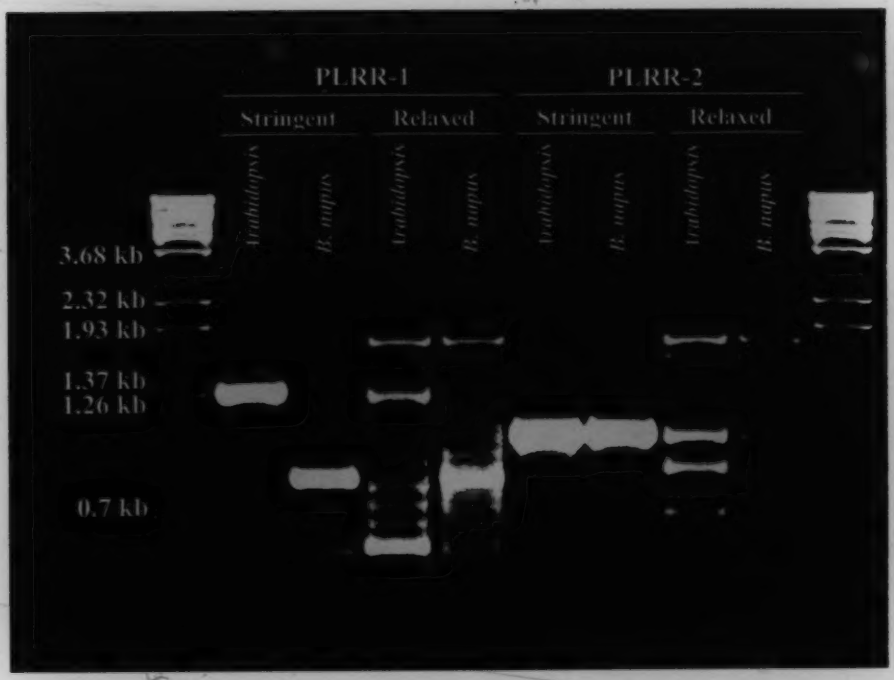


Figure 2.1. Amplification of *Arabidopsis thaliana* and *Brassica napus* DNA, with PCR primers corresponding to the *RPS2* NBS and LRR domains, under relaxed ("B") and stringent ("A") conditions (See Table 2.2).

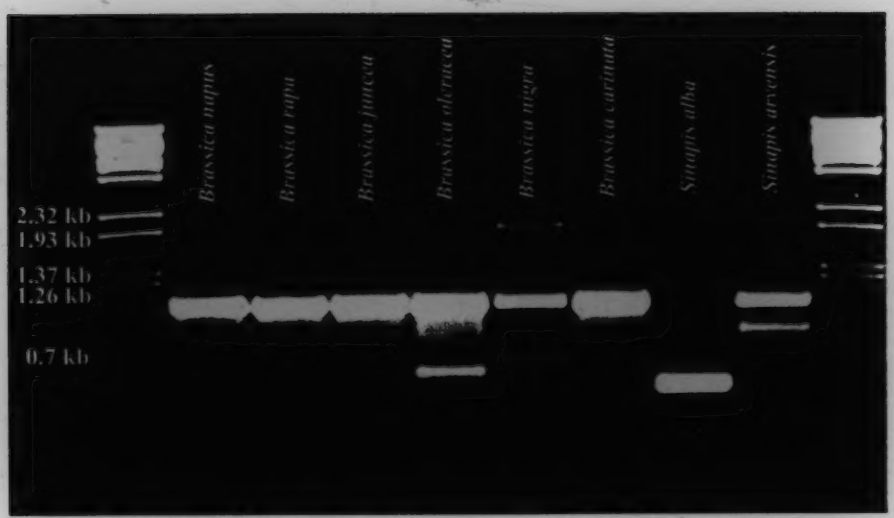


Figure 2.2. Amplification of NBS/LRR sequences in members of the Brassicaceae. PLRR-2 primers were used to amplify DNA from 8 representatives of the Brassicaceae, under stringent PCR conditions (see Table 2.2).

2.3.2 Subcloning of *Brassica napus* Resistance Gene Homologues

The *B. napus* amplification products obtained with each combination of primer pair and PCR conditions were subcloned into pBluescript KS+ (Stratagene). In all, 58 clones containing an insert were selected (see Table 2.3). Upon digestion of the plasmid DNA with *Rsa*I, clones with similar restriction patterns were grouped into families. Twenty-seven families were identified in this way. As can be seen in Figures 2.3 and 2.4, there were predominant families for each primer pair, whereas some other classes were only represented by one clone. The three main classes of clones obtained with our two primer pairs are shown in table 2.4. The restriction pattern obtained with one representative of each group is shown in Figure 2.5.

Table 2.3. Number of clones and clone families obtained from PCR experiments A, B, and C.

Total number of true positives: 58			
PLRR1:	36		
PLRR2:	22		
Total number of families: 27			
PLRR-1:	**17**	**PLRR-2:**	**10**
A :	7	A :	3
B :	5	B :	3
C :	5	C :	4

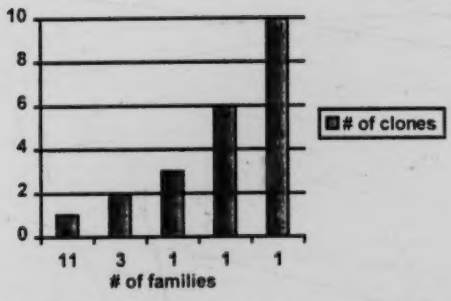


Figure 2.3. Number of clones per family, for primers PLRR-1.
Whereas ten of the clones belong to one family, eleven families were identified that were only represented by one clone.

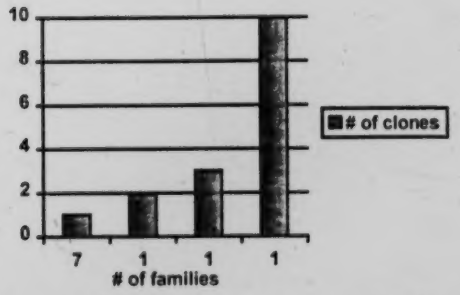


Figure 2.4. Number of clones per family, for primers PLRR-2.
Similarly to the results obtained with PLRR-1, there is one major class of PLRR-2 inserts, as well as a number of other families represented by only one or few clones.

Table 2.4. Three main classes of clones identified by their *Rsa*I restriction pattern.

Primers	Clones
PLRR1	P1-12A, P1-17A, P1-20A, P1-23A, P1-32A, **P1-3C**, P1-4C, P1-8C, P1-17C, P1-20C
PLRR1	**P1-2A**, P1-7A, P1-11A, P1-21A, P1-26A, P1-9C
PLRR2	P2-16A, P2-21A, **P2-3C**, P2-5C, P2-9C, P2-15C, P2-16C, P2-17C, P2-19C, P2-20C

The primer set used to generate each class of clones is indicated in the left column. One clone from each family, in bold, was used in the hybridization experiments.

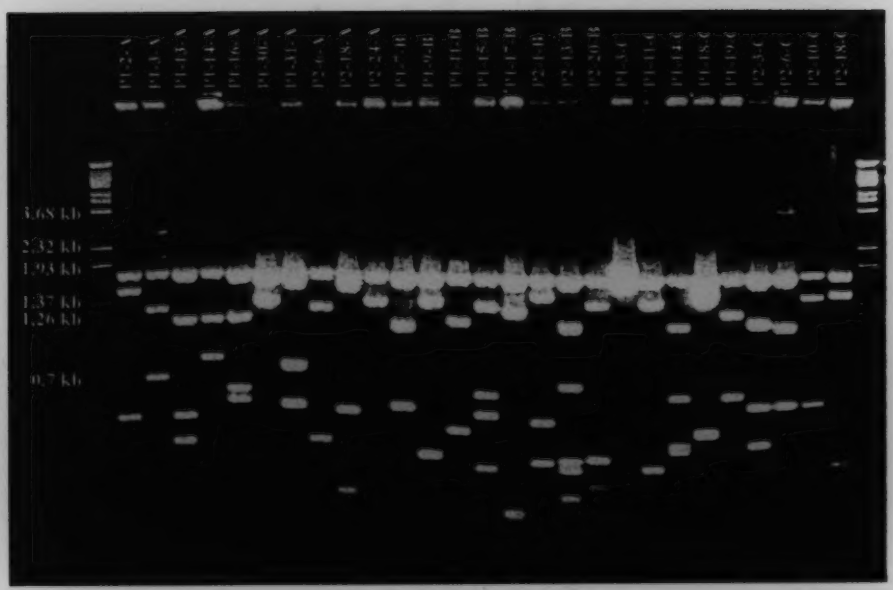


Figure 2.5. Restriction patterns obtained when DNA from one clone in each of the 27 families identified is digested with the enzyme *Rsa*I.

2.3.3 Validity of minimal mapping population

Figure 2.6 shows that, using our minimal mapping population, markers can be assigned the correct map position with a high confidence level. In this case, markers tg2h10a, tg5e11a, E32M50.90, ec2d1b, and E32M50.77, the first five markers at the top of linkage group 1 on the *B. napus* genetic map, are assigned a high LOD score (13.85, 12.64, 13.25, 13.85 and 13.85 respectively) at their respective map position, and lower scores for other positions. In addition, they show a distance above 0 cM for positions other than the correct one. Though only 5 markers and 10 map positions are shown, these observations are true for other marker pair combinations as well.

In some cases however, MAPMAKER assigns a genetic distance of 0.0 between two distant markers. This is exemplified with marker E33M49.189a (the 6th marker on the map) which, if we based ourselves on MAPMAKER's genetic distances, could map to positions 6, 7, 8, and 13 (Figure 2.7). If we look at the LOD scores for these positions however, we find that the peak is at positions 6 and 7. Using the entire mapping population, MAPMAKER assigns the same position to the 6th and 7th markers in our data set.

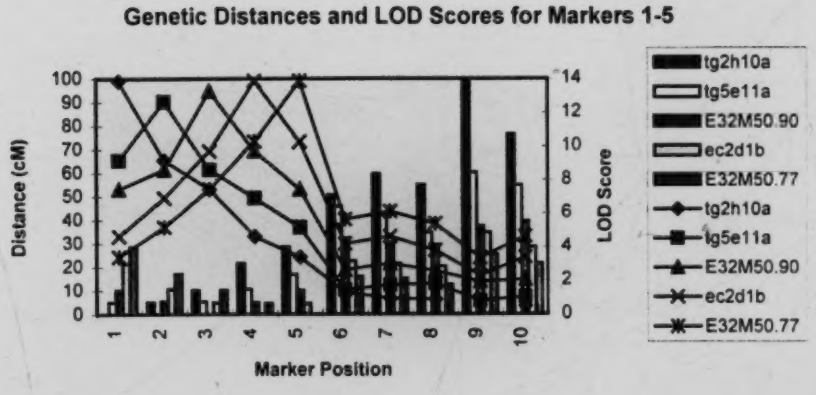


Figure 2.6. Genetic distances (▮), and LOD scores (◆) obtained with the "LOD table" command of MAPMAKER are shown for the first five markers on the *Brassica napus* genetic map.

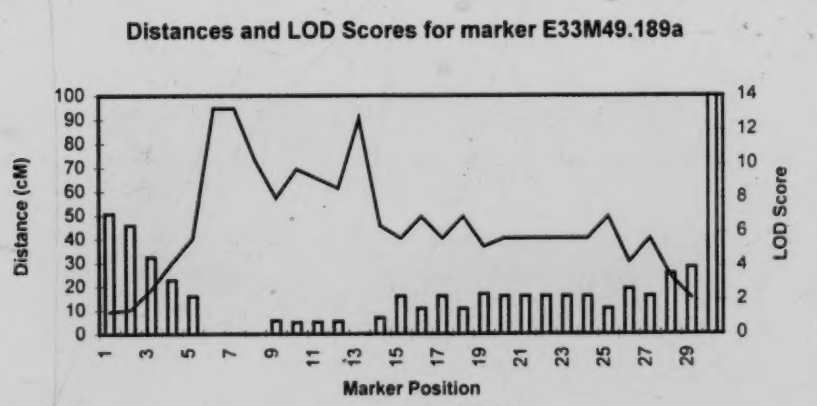


Figure 2.7. Distances (▯) and LOD scores (——) obtained with MAPMAKER's "LOD table" command are shown for E33M49.189a, the 6th marker on the *B. napus* genetic map.

2.3.4 Genetic mapping

One member from each of the 27 clone families defined by *Rsa*I restriction patterns was used as a probe on Southern blots containing genomic DNA from the *B. napus* lines Major and Stellar, digested with the enzymes *Eco*RI, *Eco*RV, *Bam*HI,

*Hind*III, *Hae*III, *Rsa*I, and *Xba*I. Twenty clones showed apparent restriction fragment length polymorphisms (RFLPs) with at least one of the enzymes used. These 20 clones could further be subdivided into 8 classes with identical hybridization patterns on genomic DNA digested with all 7 enzymes. These classes are shown in table 2.5. Theoretically, 20 classes should have been identified, based on their *Rsa*I restriction pattern. This was not the case, perhaps because clones in the same class hybridize to fragments of very similar size that are not distinguishable by this method.

Table 2.5. Grouping of polymorphic clones into 8 classes.

Class	Clones	Enzymes
1	P1-2A, **P1-3A**, P1-13A, P1-14A, P1-30A, P1-3C, P1-18C, P1-19C	*Eco*RI, *Eco*RV, *Xba*I
2	**P1-16A**, **P2-6A**, P2-3C, P2-6C, P2-10C	*Eco*RI, *Hind*III
3	**P2-18A**	*Eco*RI, *Eco*RV, *Bam*HI
4	**P1-7B**	*Eco*RV
5	**P1-15B**	*Bam*HI
6	**P2-13B**	*Eco*RI, *Hind*III
7	**P2-20B**	*Eco*RI, *Eco*RV, *Xba*I
8	**P1-11C**, **P1-9B**	*Eco*RV

The enzymes that gave polymorphisms for each class of clones are indicated. Clones that were investigated further are shown in bold.

For at least one clone from each class, the insert was amplified using the T3 and T7 primers and labeled with ^{32}P. Figure 2.8 shows an agarose gel containing a sample from each insert. The labeled inserts were then used as probes on Southern blots containing genomic DNA from the two parental lines and all 23 lines of the minimal mapping population, digested with the appropriate enzyme. A sample Southern blot of the mapping population digested with *Bam*HI, and probed with clone P1-15B, is shown in figure 2.9. Of the 8 classes, 6 showed clear polymorphisms between individuals of the mapping population, whereas 2 classes of polymorphisms could not be reproduced in the mapping population. For each polymorphism, the alleles were scored for each individual

as being the Major (A) allele or the Stellar (B) allele. As can be seen in figure 2.9, line number 346 was heterozygous at the P1-15B locus and was therefore scored as a 0 and not included in the linkage analysis. The scorings for the polymorphisms of all 6 classes are shown in table 2.6. These scorings were compared to the scorings of the other markers on the genetic map of *B. napus*, using the computer program MAPMAKER, and the most likely map position(s) was assigned to each marker (see page 83, section 2.3.6).

Figure 2.8. Agarose gel showing the inserts of the ten polymorphic clones used in our mapping studies.

Figure 2.9. Southern blot of genomic DNA from the *B. napus* parental lines Major and Stellar, digested with BamHI, and hybridized with a radio-labeled probe of the insert of clone P1-15B. The sizes in kilo base pairs (kb) of the two polymorphic bands are indicated. The numbers at the top indicate the doubled haploid line numbers. (The blots for the other clones mapped in this study can be obtained via Dr. Reinhold Mayerhofer in G404, Biological Sciences, University of Alberta.)

Table 2.6. Genotypes of the minimal mapping population for the six polymorphic clones.

Clone	Enzyme[a]	Genotypes of the Minimal Mapping Population[b]
P2-18A	EcoRV	B A A A B B A B B A B A B A B B A A A A B B B
P1-7B	EcoRV	B B(B)B B B A B B B B A B A(B)B A B B B(B)B B
P2-13B	HindIII a	A A(B)A A A A B B B B A A A A A A A A A A A A
P2-13B	HindIII b	A A B B A A B A B B B B B B B B A A B B A A A
P2-13B	HindIII d	A B B A A B A A B B(B)A A A B B A B A B B B B
P2-13B	HindIII e	A A A A A B A A B A A A A A A A A A A A A A A
P2-13B	EcoRI	A B B B(B)B B B B B(B)B B B A B B B A B A B A B
P1-15B	BamHI	A A A B B B A B A A O A A A B A A A A B B A A
P1-3A	EcoRV	A(A)A A B A B B B A A B B A B A B A B A A A A
P1-3A	EcoRI	B(A)A B A B A B B A B A B A A A A A A A B A A
P2-20B	EcoRI	B(A)O A B A B B B A A B B B(B)A B B B B A A B
P2-20B	EcoRV	B A A A B A B B B A A B B B B A B B B B A A B
P2-20B	XbaI a	A B B A A B A A B B A A A A A B B A A B B B B
P2-20B	XbaI b	A B B B A B B B B B B B A A A B B B A B B B B

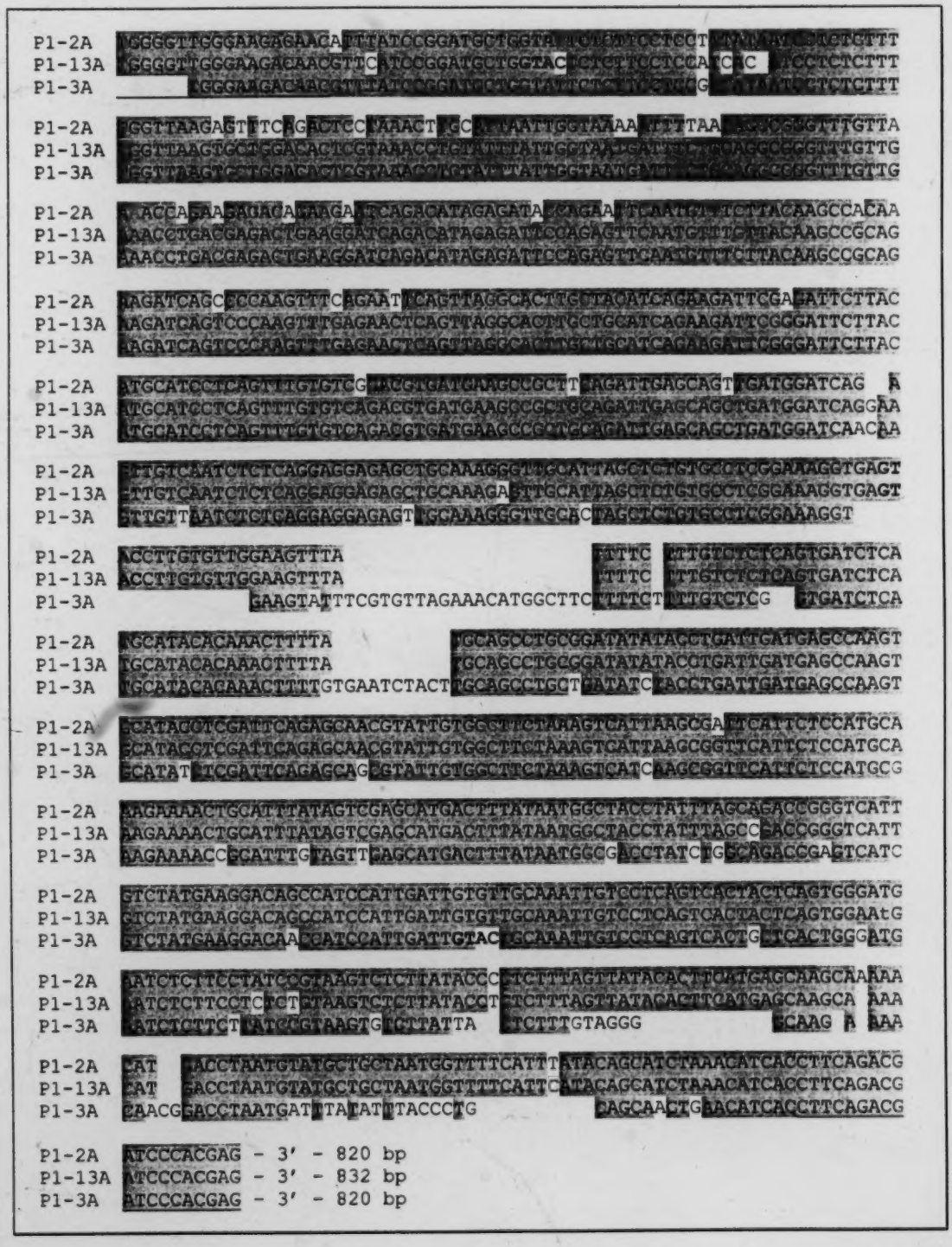
[a] When more than one polymorphism was identified for a particular enzyme, each polymorphism was assigned a lowercase letter. Hereafter, the markers are named by the primer pair (P1 or P2), the clone number, the reaction conditions (A, B, or C), the first letter of the enzyme that gave a polymorphism, and the lower case letter if applicable. eg. P2-20B-Xa.

[b] The order of the doubled-haploid lines in the minimal mapping population is: 302, 311, 312, 314, 319, 321, 331, 338, 340, 343, 346, 364, 365, 367, 370, 386, 389, 395, 398, 402, 407, and 425. The genotypes are A for Major, and B for Stellar. In the case of uncertain scorings, either the most likely scoring was indicated in parentheses and used in the mapping analysis, or the genotype was indicated as a 0 and did not contribute to the mapping of the corresponding marker. P2-20B-*Eco*RI and *Eco*RV identified the same locus.

2.3.5 Sequence Analysis

The inserts from the eight clones that were polymorphic between the plants Major and Stellar (P1-3A, P1-16A, P2-18A, P1-7B, P1-15B, P2-13B, P2-20B, and P1-11C), as well as clones P1-2A and P1-13A, which are in the same family as P1-3A, were sequenced using the T3 and T7 primers of the pBluescript vector. The sequences from both strands of each clone were then analyzed using the program SeqAid™ II version 3.5, as to find overlaps between the sequences, which would indicate that the whole insert had been sequenced. Such alignments were found for seven out of the ten clones. The inserts from P1-16A, P1-15B, and P2-13B, estimated to be approximately 1.69 kb, 2.2 kb, and 1.9 kb in size respectively (see figure 2.8), were not sequenced completely in this first round. All sequences are reported below in figures 2.10 to 2.17. As P1-2A, P1-3A and P1-13A gave the same hybridization pattern, their sequences were aligned using SeqAid. This alignment is shown in figure 2.10.

Figure 2.10. DNA sequence alignment of clones P1-2A, P1-3A, and P1-13A. Identical nucleotides are shaded. *Rsa*I restriction sites are in bold. Primer sequences are underlined.


P1-2A
P1-13A
P1-3A
P1-2A - 3′ - 820 bp
P1-13A - 3′ - 832 bp
P1-3A - 3′ - 820 bp

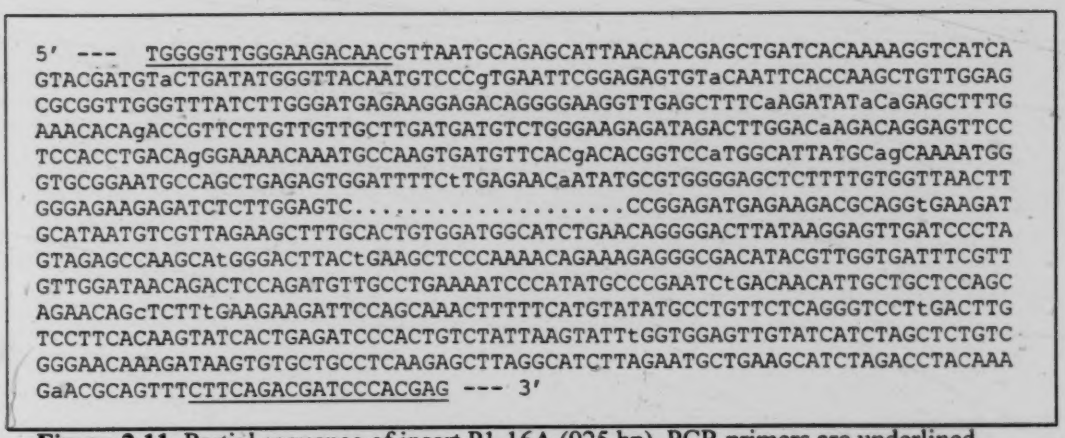


Figure 2.11. Partial sequence of insert P1-16A (925 bp). PCR primers are underlined. Uncertainties in the sequence are indicated by lowercase letters. The gap in the sequence refers to the portion that has not yet been sequenced.

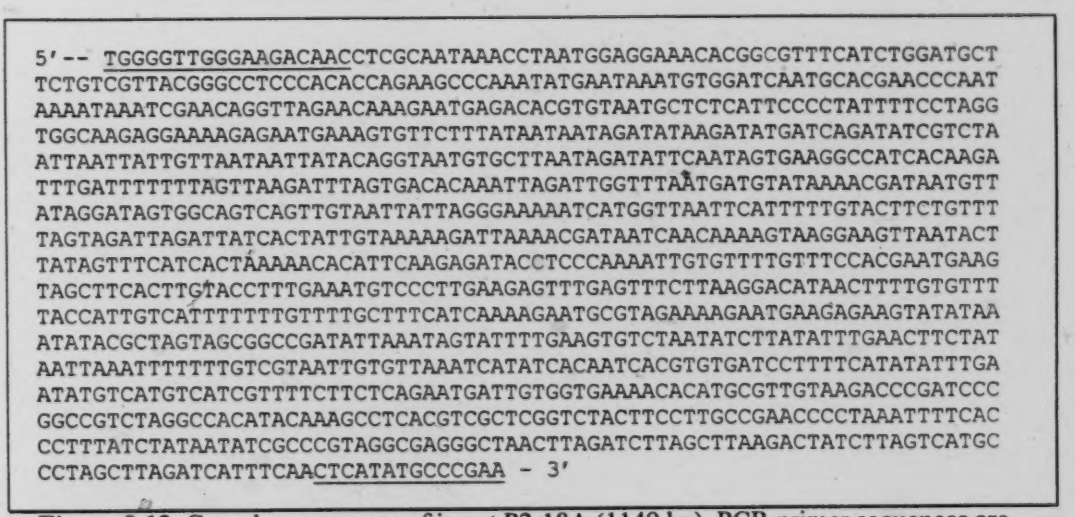


Figure 2.12. Complete sequence of insert P2-18A (1149 bp). PCR primer sequences are underlined.

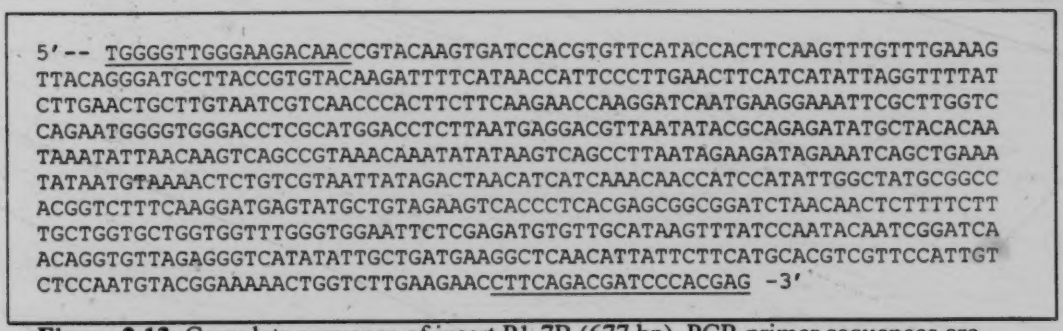


Figure 2.13. Complete sequence of insert P1-7B (677 bp). PCR primer sequences are underlined.

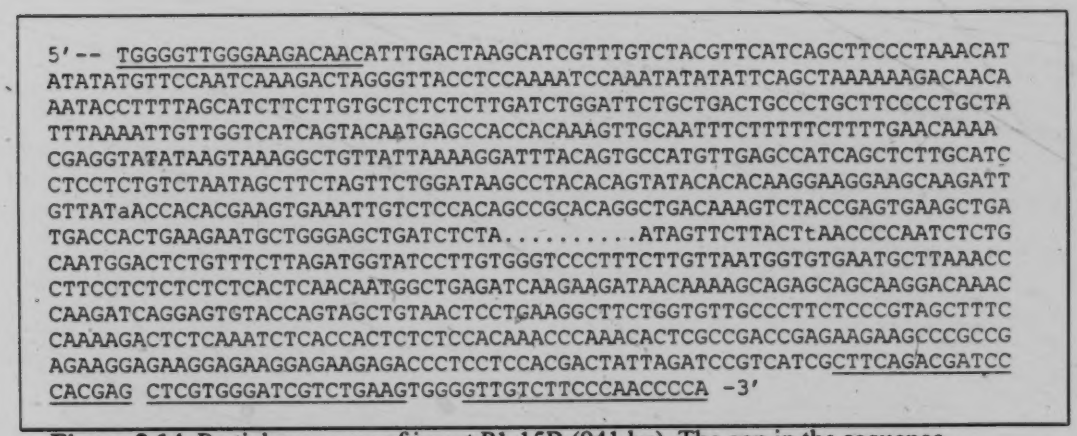


Figure 2.14. Partial sequence of insert P1-15B (941 bp). The gap in the sequence represents the portion that has yet to be sequenced. Uncertainties in the sequence are indicated by lowercase letters. The 3' end of the sequence seems to contain another insert, as evidenced by the presence of two extra primer sequences. The PCR primer sequences are underlined.

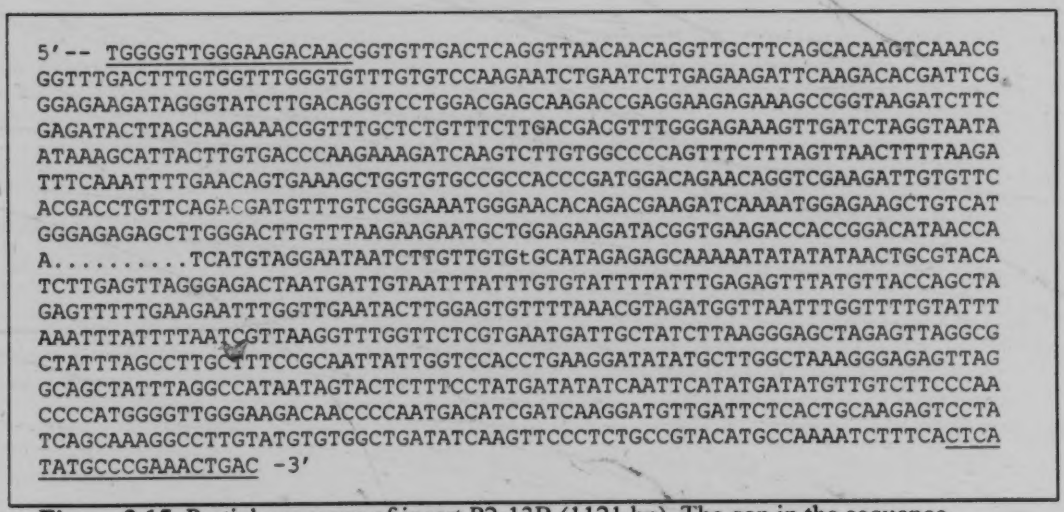


Figure 2.15. Partial sequence of insert P2-13B (1121 bp). The gap in the sequence represents the portion that has yet to be sequenced. Uncertainties in the sequence are indicated by lowercase letters. The PCR primer sequences are underlined.

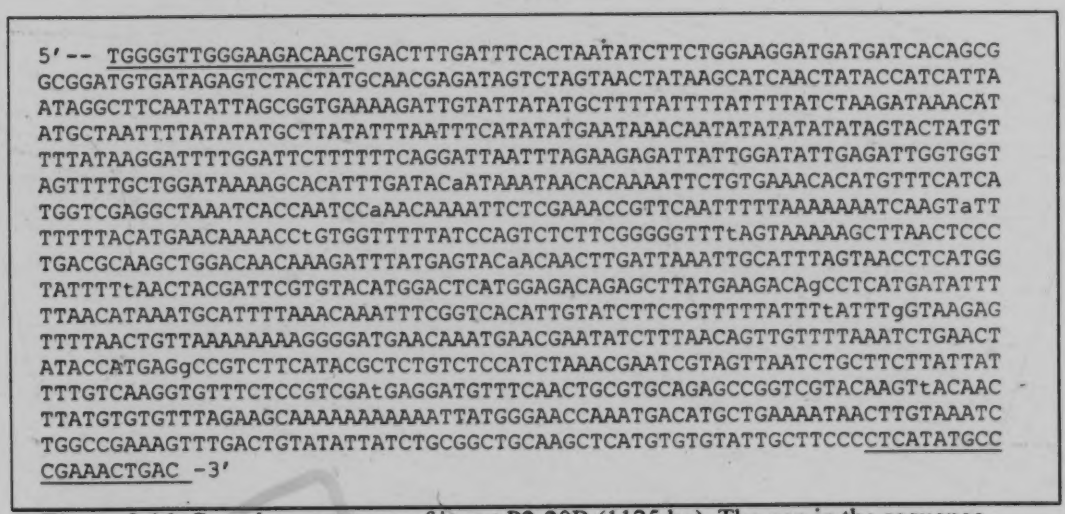


Figure 2.16. Complete sequence of insert P2-20B (1125 bp). The gap in the sequence represents the portion that has yet to be sequenced. Uncertainties in the sequence are indicated by lowercase letters. The PCR primer sequences are underlined.

5'-- TGGGGTTGGGAAGACAACATTTGACTAAGCATCATTTGTCTACGTTCATCAGCTTCCCTAAACAT
ATATATATATGTTCCAATCAAAGACTAGGGTCAGCTCAAAATCCAAATTTATATTCAGCTCAAATGACAA
CAAAATACCTTTTAGCATCTTCTTGTGCTCTCTCTTGATCTGGATTCTGCTGACTGCCTTGCTTCCCCTG
CTGCTATTTATATTGTTCAGTCATCAGTACAATAAGCCACCACAAAGTCTGTATGTACAGTTCTTTTTTT
CTTAAAACAAAACAGGTATATAAGTGAAGGTTGTTATTAAAAGGATTTACAGTGCCATGTTGAGCCATGA
GCTCTTGCATCCTCCTCTGTCTAATAGCTTCTAGTTCAGGATCAGCCTACACAAGGAAAGGAAGTAACAG
AAGCAAACTTTTAACATGATGCAGTTACAGTGCCATGCCATGATCGTTCAAAAAGAGTCCATTCTTTTTC
TTTTTCATTTAACACTGAACATAACTCATTTGTTCTTAATAGATCTCGACTTCAGACGATCCCACGAG-3'

Figure 2.17. Complete sequence of insert P1-11C (553 bp). The PCR primer sequences are underlined.

The sequences were translated in all six reading frames and compared to known protein sequences via a Gapped Blast search (Altschul *et al.*, 1997). The main homologies found, as well as the corresponding Expected (E) values, which are the number of similar matches to be expected by chance alone, are reported in table 2.7. Other than for P2-18A, significant homologies were found for all the sequences. Clones P1-7B, P1-15B and P1-11C shared significant homology with *Arabidopsis* and rice coding sequences of unknown function. Clone P1-15B also has weak homology to a rice resistance gene homologue.

A *B. napus RPS2* homologue has been found in clone P1-16A. In fact, over the length of P1-16A, there was 90% identity to *RPS2* at the nucleotide level. This clone however was monomorphic in our mapping population, and was not assigned a map position. The 5' portion of the P2-13B clone encompassing the conserved NBS domain was also highly homologous to several RPS2-like sequences, as well as to RPS2 itself. The alignments of RPS2 with P1-16A and P2-13B is shown in figure 2.18. The major class of clones, represented by P1-2A, P1-3A, and P1-13A, corresponded to human and bacterial RNAase L inhibitors, a probable yeast membrane protein, and a hypothetical ATP-Binding Cassette (ATP) transporter. Human RNAse L inhibitor is a P-loop-

containing protein involved in the interferon pathway, and possibly in general RNA stability (Bisbal *et al.*, 1995). ATP transporters confer resistance to a group of toxic compounds in bacteria, protozoans, fungi and mammals (reported in Alarco *et al.*, 1997). In particular, they can confer cadmium resistance in humans and yeast (Tommasini *et al.*, 1996). *Arabidopsis* ATP transporters were found to be induced upon toxin treatment (Tommasini *et al.*, 1997), including cadmium, and they seem to be induced in response to powdery mildew infection (S. Somerville, personal communication). The plant pathogen *Agrobacterium tumefaciens* also requires ATP transporters for virulence in its host (Matthysse *et al.*, 1996; Kemner *et al.*, 1997). Finally, an ATP transporter mutation was recently implicated in a human retinal disorder (Allikmets *et al.*, 1997).

Another clone, P2-20B, showed low DNA sequence homology to the 5' flanking region of human glutathione S-transferase (GST). Such proteins are also found in plants and are thought to be involved in neutralizing hydrophobic, cytotoxic compounds such as herbicides, and possibly in limiting the cellular damage caused by reactive oxygen species during the defense response (Chen *et al.*, 1996). In fact, transcription of plant GST genes can be induced by H_2O_2 (Chen *et al.*, 1996; Tenhaken *et al.*, 1995) and heavy metals such as cadmium (Marrs and Walbot, 1997). Interestingly, *Arabidopsis* ATP transporters are also induced upon treatment with herbicides (Tommasini *et al.*, 1997) and at least one of them, *AtMRP1*, seems to function in conjunction with a glutathione transferase, as it transports glutathione S-conjugates of toxic compounds out of the cytoplasm (Lu *et al.*, 1997). In this respect, it is noteworthy that, in yeast, transcription factors homologous to the mammalian Fos and Jun factors induce both ATP transporters and genes involved in oxidative stress responses (Alarco *et al.*, 1997).

Table 2.7. Homology search results for ten of the NBS/LRR sequences identified

Clones	Homologies	E value[a]
P1-2-A	Human RNAase L inhibitor	$1x10^{-63}$
	Probable yeast membrane protein	$9x10^{-55}$
	Hypothetical ABC transporter	$7x10^{-30}$
	Methanobacterium RNAase L inhibitor	$2x10^{-26}$
P1-3-A	(same as P1-2-A)	--
P1-13-A	(same as P1-2-A)	--
P1-16-A	*Arabidopsis* resistance gene *RPS2*	$7x10^{-64}$
	Arabidopsis, rice and tomato R gene homologues	$2x10^{-4} - 3x10^{-27}$
P2-18-A	No significant homology found	--
P1-7-B	Hypothetical Protein (*A. thaliana*)	$3x10^{-25}$
P1-15-B	Putative rice disease resistance gene	1.1
	Rice and *Arabidopsis* anonymous cDNAs	$1x10^{-9} - 7x10^{-16}$
P2-13-B	*Arabidopsis*, rice and tomato R gene homologues	$0.018 - 1x10^{-29}$
	Arabidopsis resistance gene *RPS2*	$3x10^{-21}$
P2-20-B	Human glutathione S-transferase 5' flanking region	0.00027
P1-11-C	Rice and *Arabidopsis* anonymous cDNAs	$2x10^{-6} - 5x10^{-14}$

[a] The E (expected) value refers to the number of similar matches expected by chance alone. The lower the E value, the more significant the match is. The E values reported here for each sequence are the ones obtained by performing a Gapped Blast search (see text for details.

A. NBS domain

```
P1-16A 1    GVGKTTLMQSINNELIT-KGHQYDVLIWVTMSREFGECTIHQAVGARLGL---SWDEKET
            GVGKTTLMQSINNELIT KGHQYDVLIWV MSREFGECTI QAVGARLGL   SWDEKET
RPS2   185  GVGKTTLMQSINNELIT-KGHQYDVLIWVQMSREFGECTIQQAVGARLGL---SWDEKET
            GVGKTT++  +NN L+  K + +D ++WV +S+      IQ  +  ++G    SW  K T
P2-13B 1    GVGKTTVLTQVNNRLLQHKSNGFDFVVWVFVSKNLNLEKIQDTIREKIGYLDRSWTSK-T

P1-16A      GEG*AFKIYRALKHRPFLLLLDDVWEEIDLD----------------------------
            GE  A KIYRAL+ + FLLLLDDVWEEIDL+----------------------------
RPS2        GENRALKIYRALRQKRFLLLLDDVWEEIDLE----------------------------
             E +A KI+  L +KRF L LDDVWE++DL
P2-13B      EEEKAGKIFEILSKKRFALFLDDVWEKVDLGNNKALLVTQERSSLVAPVSLVNF*DFKF*

P1-16A      --KTGVPPPDRENKCQVMFTTRSMALCSKMGAECQMRVDFLENNMRGELFC 136
            --KTGVP PDRENKC+VMFTTRS+ALC+ MGAE ++RV+FLE     ELFC
RPS2        --KTGVPRPDRENKCKVMFTTRSIALCNNMGAEYKLRVEFLEKKHAWELFC 320
              K GVP PD +N+ K++FTT S  +C  MG + K+++E L  + AW+LF
P2-13B      TVKAGVPPPDGQNRSKIVFTTCSDDVCREMGTQTKIKMEKLSWERAWDLF  169
```

B. LRR Domain

```
Consensus     PXXαXX LXXLXXLXαXXXX αXXα
P1-16A   ?    PKTERG RHTLVISLLD NR LQML
              PK E   R  LVISLLD NR +Q L
RPS2     505  PKAENW RQALVISLLD NR IQTL

P1-16A        PENPIC P               I
              PE  IC P               I
RPS2          PEKLIC PK LTTLMLQQNSSLKKI

P1-16A        PANFFMYMPVLRVLDLSFTS ITEI
              P  FFM+MPVLRVLDLSFTS ITEI
RPS2          PTGFFMHMPVLRVLDLSFTS ITEI

P1-16A        PLSIKY LVELYHLALSGTK ISVL
              PLSIKY LVELYHL++SGTK ISVL
RPS2          PLSIKY LVELYHLSMSGTK ISVL

P1-16A        PQELRH LRMMKHLDLQRTQFLQTI PR
              PQEL + LR +KHLDLQRTQFLQTI PR
RPS2          PQELGN LRKLKHLDLQRTQFLQTI PR
```

Figure 2.18. Amino acid sequence of conserved R gene domains, for *RPS2* and the clones P1-16A and P2-13B. Identical amino acids are indicated by the corresponding letters, and conservative substitutions are indicated by (+). **A**. The sequence encompassing the first two nucleotide-binding site (NBS) conserved domains (kinase 1a and 2a) is shown, with these two domains shaded. Gaps are indicated by (-). Stop codons are indicated by an asterisk (*). The sequence of clones P1-16A and P2-13B was not complete, so it was not possible to demonstrate conservation of the third conserved NBS domain. **B**. The amino acid sequence of RPS2's first 5 leucine-rich repeats (LRRs) is shown aligned with P1-16A. The consensus sequence is shown, with X being any amino acid, and α being aliphatic amino acids (L/I/V). Amino acids that fit this consensus are shaded. The *RPS2* sequence and the nature of the conserved domains were obtained from Mindrinos *et al.* (1994).

2.3.6 Quick mapping using the polymerase chain reaction

Fifteen polymorphisms were found when DNA from the two parental lines was amplified with both our primer pairs and two sets of reaction conditions, and digested with eight restriction enzymes. A sample agarose gel showing a polymorphism between the parental lines is shown in figure 2.19. DNA from the individuals in the minimal mapping population was then amplified with the same primers and reaction conditions that gave polymorphisms, and was then digested with the appropriate enzyme. This can be seen in figure 2.20 for the PLRR-1/Stringent/*Rsa*I combination. Once again, the polymorphic bands were scored as Mendelian traits originating from Major or Stellar. Though some of these polymorphisms were redundant, we were able to map at least 4 loci in a very short amount of time. In fact, it is possible to go from DNA isolation to map position in just two days. The genotypes of the minimal mapping population for the four loci mapped are shown in table 2.8. These loci were assigned a map position on the *B. napus* genetic map as described above for the RFLP markers. Many other loci could probably be identified with our two primer pairs.

The genetic map showing all the markers mapped in this study, as well as the flanking markers, is shown in Figure 2.21. It is interesting to note that three of the PCR markers were closely linked to the RFLP markers, though there was recombination between all the markers and none mapped exactly to the same position. The markers mapped both singly and as clusters. The largest cluster was on linkage group (LG) 9, which also contains the white rust resistance locus *Acal*. Two markers, P2-13B-Hb and P2-13B-Ha, also mapped in the vicinity of the pathogenesis-related gene *Pr2*. The position of P2-13Hb, however, is uncertain, as we were unable to distinguish between

three possible loci. All three locations are reported, and possible reasons are discussed in section 2.4.

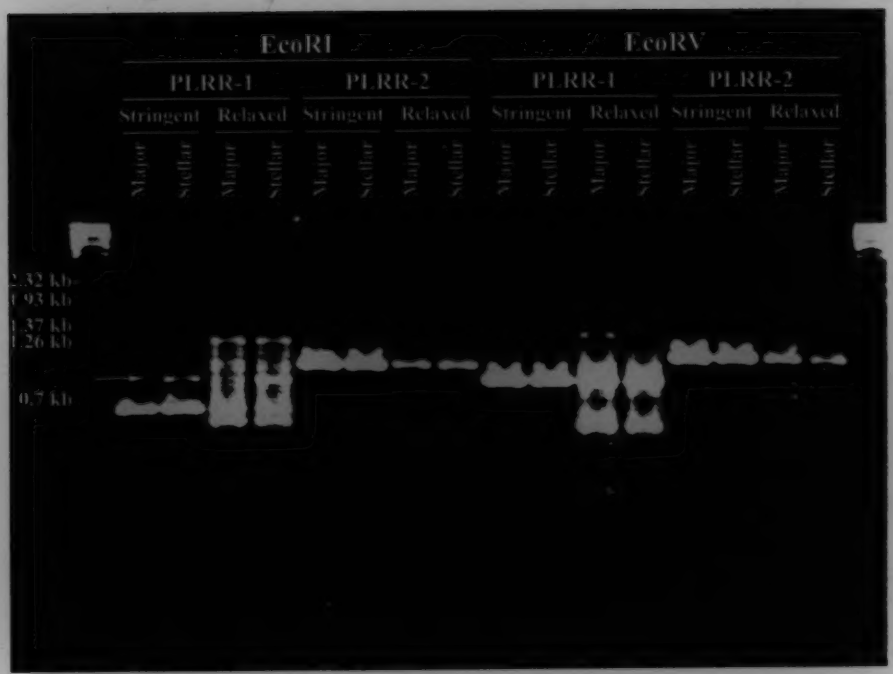


Figure 2.19. DNA amplification of the parental lines Major and Stellar, using primer sets PLRR-1 and PLRR-2 under two sets of reaction conditions (stringent vs. relaxed), and digestion with the enzymes EcoRI and EcoRV. The arrow indicates one of the polymorphisms that can be seen between the parental lines.

Figure 2.20. DNA amplification of the minimal mapping population with primers PLRR-1 under stringent conditions, and digestion with RsaI. The arrow indicates the polymorphism in the population. The DH line numbers are indicated at the top.

Table 2.8. Genotypes of the minimal mapping population for PCR-generated markers.

Polymorphisms[a]	Genotypes of the Minimal Mapping Population[b]																					
P1-S-RsaI	B	A	A	B	A	B	A	B	B	A	B	A	B	A	A	A	A	A	A	A	B	A	B
P1-R-HaeIII	A	B	B	A	B	B	A	B	B	B	A	A	B	B	B	B	A	B	A	B	B	A	B
P1-R-SauA	A	B	B	B	A	B	A	A	B	B	B	A	B	A	B	B	A	B	A	B	B	B	B
P1-R-SauB	A	B	A	B	B	B	B	B	B	B	B	B	B	A	B	A	B	A	B	A	A	A	A

[a] With the enzyme Sau3A, two polymorphisms were identified. This is indicated with an upper case letter after the enzyme name. The markers are named by the primer pair (P1), the reaction conditions (S for stringent, or R for relaxed), the restriction enzyme that gave a polymorphism, and the upper case letter if applicable.

[b] The order of the doubled-haploid lines in the minimal mapping population is: 302, 311, 312, 314, 319, 321, 331, 338, 340, 343, 346, 364, 365, 367, 370, 386, 389, 395, 398, 402, 407, and 425. The genotypes are A for Major, and B for Stellar.

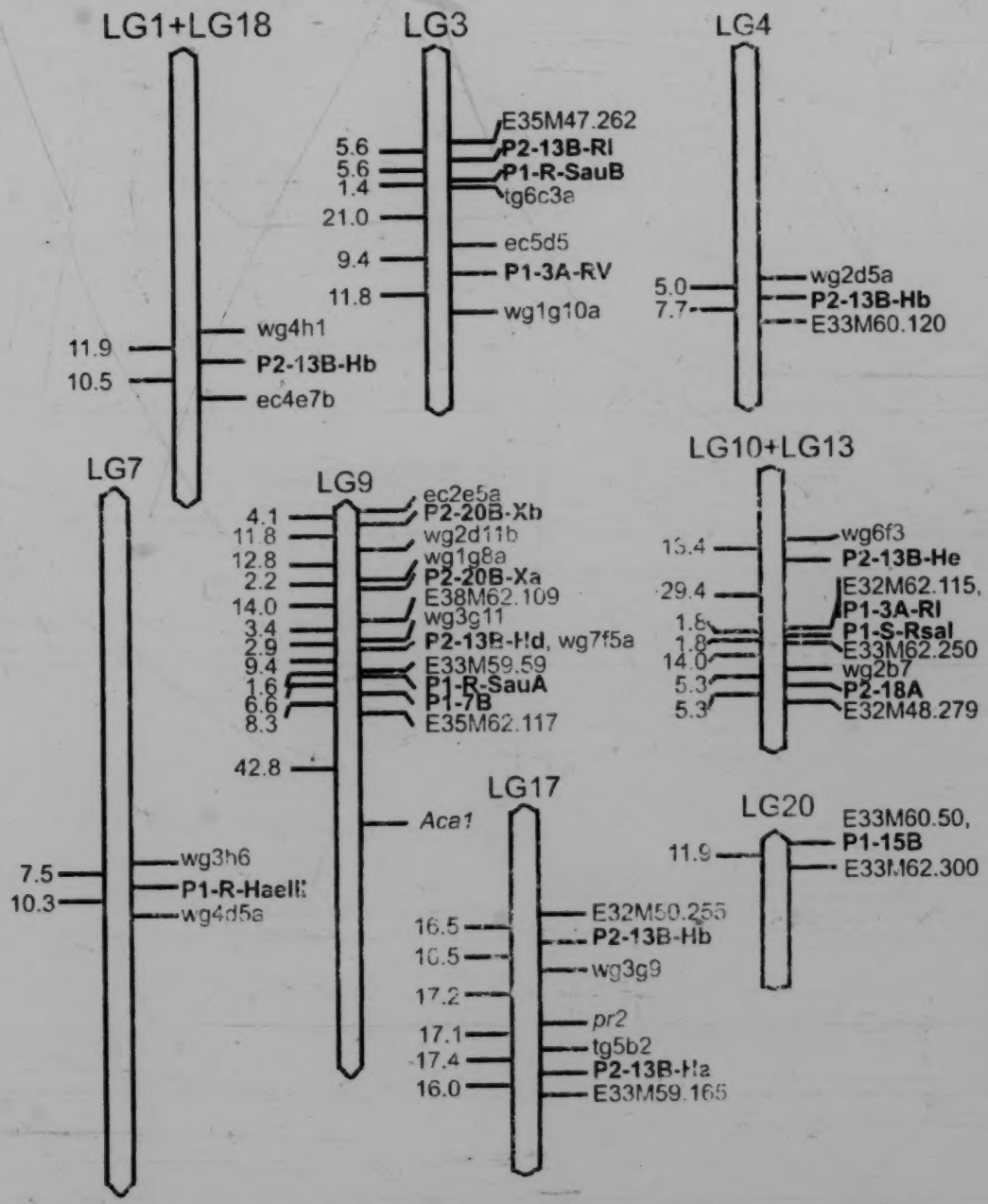


Figure 2.21. *Brassica napus* genetic map showing 8 of the 19 linkage groups, as well as the map position of the loci identified in this study and their flanking markers. Numbers on the left of each linkage group (LG) are genetic distances, expressed in centimorgans (cM). The markers identified in this study are shown in bold.

2.4 Discussion

We have shown that PCR primers corresponding to conserved disease resistance gene domains could amplify DNA from closely, as well as distantly related species of the Brassicaceae. This primer set could therefore be used to amplify and map sequences, similar to the ones we present here for *Brassica napus*, in other species of this family. This includes *B. rapa* and *B. juncea*, which are both presently receiving a considerable amount of interest from plant breeders. Our results show that some amplification products are of identical size between different species. However, some are unique to particular species. If these unique sequences are also resistance gene homologues, then there may be a benefit in looking at different species, as this will allow us to increase our supply of potential resistance genes. As resistance genes have often been lost in the crop domestication process, and as breeders frequently have recourse to wild species for resistance genes, having information on the genetic position of resistance loci in wild relatives of useful crops would be valuable.

Based on the *Rsa*I restriction pattern of cloned P-loop/LRR amplification products, one family was identified on average for every two clones (table 2.3). This does not accurately represent our results however, as we have found that, in each case, most clones were grouped into the same family (table 2.4), whereas a few other clones were in a class by themselves. This may only indicate that not all PCR products have an equal probability of being cloned. Alternatively, it may indicate that each primer pair preferentially amplifies certain classes of sequences, which would mean that different primer pairs must be used to identify all possible classes of resistance gene homologues. This is supported by the results of Leister *et al.* (1996), which show that primers differing

by as little as 2 nucleotides allowed the identification of two different groups of sequences. Our results also show that varying the PCR reaction conditions further increases the number of sequences identified. Clone P2-13B, for example, was generated by using relaxed PCR conditions, is in a class by itself, and is highly homologous to other resistance gene homologues.

Though different clone families had different restriction patterns, they could be grouped into larger classes based on their hybridization patterns on *B. napus* DNA. That is the case for P1-2A, P1-3A, and P1-13A which appear to map to the same locus based on their hybridization pattern, but are slightly different at the DNA sequence level, generating different *Rsa*I restriction patterns. These differences may however be partly due to mistakes made by the Taq polymerase in the initial amplification of these sequences. Confirmation for this would have to come from the sequencing of corresponding genomic clones. Nevertheless, these results suggest that both selection processes are important, as a classification based on hybridization alone would not have allowed us to distinguish between individual members of the same family. This was also shown by Wang *et al.* (1996). In their study, seven *Arabidopsis* BAC clones were found to hybridize to a probe corresponding to a resistance gene homologue. These clones however could be subdivided into 4 distinct classes, based on their restriction pattern with *Hind*III.

We used doubled-haploid lines of *Brassica napus* to map our P-loop/LRR sequences. Such lines greatly simplify mapping studies, as they are the products of single meiotic events. The segregation of only two parental alleles, instead of four, is analyzed at a time (Raeder and Broda, 1986). Some complications arise, however, in cases such as

P1-15-B, where one of the individuals appears to be an heterozygote for the locus of interest (see figure 2.9). This is probably due to a nonreciprocal chromosomal translocation, resulting in the loss of one allele in one individual, and the gain of that same allele in an other. Such translocation events are common in the genome of *B.napus* (Sharpe *et al.*, 1995).

Except for clone P1-16A, all our inserts hybridized to multiple loci, many of which were monomorphic, and some of which were polymorphic in our mapping population. We found that different polymorphic loci, identified by the same probe, could map to very different positions on the genetic map. This was the case for P2-13B. It identified 4 distinct polymorphic loci when it was hybridized to genomic DNA of the mapping population, digested with *Hind*III (see table 2.6). Locus P2-13B-Ha maps to the bottom of LG17, P2-13B-Hd maps to the upper third of LG9, and P2-13B-He is located at the top of LG10+LG13 (see figure 2.21). The position of P2-13B-Hb could not be resolved, but the three possible positions for this marker were distinct from that of the other three P2-13B-H loci. On the other hand, some probes identified multiple loci that map to similar locations. Marker P2-20B-X, for example, identified 2 loci that mapped to the top of LG9. Finally, different probes recognized loci that were clustered together, such as those at the top of LG9. Note that the white rust resistance gene *AcaI* also maps to LG9, albeit 51 cM away from our closest marker.

Our results therefore show that P-loop/LRR probes can identify loci throughout the genome, and that these loci are found both separately and in clusters. This was also found to be true in similar studies with soybeans (Kanazin *et al.*, 1996) and potatoes (Leister *et al.*, 1996).

Nine of the ten groups of sequences that we have identified using this PCR approach have significant homology to known sequences. We have identified a *B. napus RPS2* homologue (P1-16A), as well as a NBS-containing sequence (P2-13B) that is also homologous to *RPS2*. Interestingly, P1-16A lacks part of the second leucine-rich repeat found in *RPS2* (see figure 2.18). If LRR domains are important specificity determinants in resistance, as is suggested by studies with the *M* gene (Anderson *et al.*, 1997), then P1-16A may correspond to a type of *RPS2*-like resistance gene with a different resistance specificity. Alternatively, it may correspond to an allele that has lost the specificity encoded by the full RPS2 homologue in *B. napus*. P2-13B appears to lack a LRR domain altogether, although it is difficult to make this conclusion based on our results, as reverse primer number 2 lies at the beginning of the second leucine-rich repeat in *RPS2*. Thus we can only say that P2-13B lacks the first and part of the second leucine-rich repeat. So far, all the R genes that have been found to possess a NBS domain also possess a LRR domain. There are several possible explanations for our results. P2-13B could be a novel type of R gene that possesses a nucleotide-binding site but no LRR domain. It seems however that LRR proteins are essential in resistance, as evidenced by the need for the LRR-containing *Prf* gene in *Pto*-mediated resistance (Salmeron *et al.*, 1996). We could therefore speculate that a protein homologous to *Prf* acts in concert with the P2-13B gene product. Another possibility is that P2-13B possesses a LRR domain that is located in a different part of the sequence. This will only be determined once the complete sequence of this clone is known and/or when a corresponding cDNA is sequenced. It is also possible that P2-13B corresponds to a NBS-containing gene that does not function in resistance. The homology between resistance gene homologues and the N-terminal

portion of the predicted P2-13B peptide is not restricted to the conserved NBS domains however. Finally, it is possible that P2-13B corresponds to a pseudo-gene that has lost its LRR domain through unequal recombination. This is supported by the fact that numerous stop codons are found in the second half of the P2-13B sequence. Interestingly, the predicted amino-acid sequence of this second half is relatively leucine/isoleucine-rich and contains a number of prolines, all characteristic of LRR domains.

A second group of sequences, P1-2A, P1-3A, and P1-13A, was found to have strong homology to a human RNAse L inhibitor and to various ATP transporters. Both gene types are involved in the disease response at some level. Human RNAase L inhibitor is central to the interferon pathway, which is involved in activating the immune response (Bisbal *et al.*, 1995). ATP transporters seem to have multiple roles, but they are known to be involved in human (Allikmets *et al.*, 1997) and plant disease resistance (S. Somerville, personal communication). Interestingly, ATP transporters also depend on an interferon-like pathway, at least in yeast, as they are induced by transcription factors that are homologous to AP-1 (Alarco *et al.*, 1997). The AP-1 transcription factor is formed of Fos and Jun subunits, the expression of which is induced by interferon-γ (Rubio, 1997). This analogy may or may not be significant, but it is reasonable to think that proteins that act in response to similar stimuli may share at least some homology. If this were true, then the power of gene identification by homology would be greatly increased, as it would mean that by looking for homologues of a particular gene one could find other genes that are involved in the same response. In this respect, it is interesting that we have found two types of sequences that appear to be involved at different stages of disease resistance while looking only for the receptors.

We have also found a sequence with very low homology (65% identical over 94 nucleotides) to an intriguing AT-rich section in the 5' flanking region of human glutathione S-transferase, which is also regulated by AP-1 transcription factors (Morrow *et al.*, 1989). This region of homology did not encompass the AP-1 binding site however. This finding is not likely to be significant, as far as our search for resistance gene homologues is concerned, nonetheless, the coincidence is peculiar.

Finally, we have identified three sequences, P1-7B, P1-15B, and P1-11C, which are homologous to rice and *Arabidopsis* coding sequences of unknown function. P1-15B shares very weak amino acid similarity with a rice resistance gene homologue, but this homology is not significant (E=1.1). It will be interesting, as the function of these sequences is elucidated, to see whether they play a role in the plant disease response.

We have shown that rough genetic mapping can theoretically be accomplished with considerably fewer individuals, if these are chosen to represent at least one recombination event at set intervals. Our data also show that difficulties in resolving map positions using a minimal population often reflect similar problems in the full mapping population. It should be noted however that very fine mapping will always require an increasing number of individuals. Once again, individuals can be chosen to represent additional recombination events in the desired region only. This should limit the number of individuals that one needs to add to a population to map a defined chromosomal region.

This method is most powerful when there is a good coverage of the genome with markers. It is also of crucial importance that there be a high level of confidence in the order of markers that flank the chosen intervals. This is because the importance of each

interval is artificially inflated when this method is used. If there is an error in the order of markers, then a new marker may not be assigned a position because it belongs to an interval that is not defined by this order of markers. Similarly, it may be assigned a position within a wrong interval only because it is associated with one of the interval's flanking markers. To increase the probability of determining the correct order of markers in the initial genetic map, a high number of individuals must be scored accurately for each marker. Also, the alleles of each marker must segregate randomly in the population, which is not always the case, due presumably to methylation effects (D. Lydiate, personal communication). In short, before a good minimal population can be defined, an accurate saturated genetic map must be generated by analyzing a large number of individuals with a larger number of markers. In addition, it would be beneficial to add a few randomly chosen individuals to the minimal mapping population. This would not add any information in terms of recombination events, but it would increase the confidence level with which a marker is mapped to a particular interval. This is because it would allow one to test whether it is more likely that the scorings are wrong, and/or that the assigned map position is wrong. If there is segregation distortion for a given marker, and this same distortion is found in the additional random individuals, then something is likely to be affecting the scorings for that marker. Also, if the scorings in the extra individuals match the scorings for the flanking markers of the given map position, then this position is more likely to be correct, and any discrepancies in the minimal mapping population are probably due to scoring errors.

Our results did not allow us to make such distinctions, as we used a strictly minimal population. As a result, some polymorphic markers were not assigned a map

position, and others were assigned more than one. In most cases, it was possible to choose the most likely position by looking at the genotypes of the flanking markers. The best position is probably that which minimizes the number of double recombination events between the flanking markers and our sequences. In the case of P2-13B-Hb however, we were not able to select the most likely position from the three loci suggested by MAPMAKER. All three positions require that at least three double recombination events have occurred between P2-13B-Hb and its flanking markers (figure 2.22). This is probably due to scoring errors for P2-13B-Hb and/or the flanking markers. Alternatively, it is possible that sequences that cross-hybridize with P2-13B-Hb map to different loci, and that more than one position is correct. Distinguishing between these two possibilities would require that additional individuals be genotyped for this marker.

<u>LG1+LG18</u>

wg4h1	A-BBAABAB-B-BA-AA-ABAA-
P2-13B-Hb	AABBAABABBBBBBBBAABBAAA
ec4e7b	AABBAABABBBBBABAABABAAB

<u>LG4</u>

wg2d5a	-----ABABBBB--BAABBBAAA
P2-13B-Hb	AABBAABABBBBBBBBAABBAAA
E33M60.120	AABABABABBBBABBAAABBAAA

<u>LG17</u>

E32M50.255	AAABBABBBBABBBBAABBBAAA
P2-13B-Hb	AABBAABABBBBBBBBAABBAAA
wg3g9	A-BBAAAABBAABBBAABABAAA

Figure 2.22. Genotypes of the 23 individuals of the minimal mapping population for P2-13B-Hb and its flanking markers in three possible map locations. 'A' is the Major allele, and 'B' is the Stellar allele. Probable sites of double recombination events are shaded.

Despite these problems, we have mapped a number of resistance gene homologues, as well as other sequences of unknown function, in *Brassica napus*. None of these however were linked to either blackleg or white rust resistance. Because no other resistance loci have been identified in *B. napus* so far, our sequences are of limited use at the moment, except for the fact that some point to loci that most probably contain resistance genes. We have therefore designed a PCR-based approach to quickly test whether primer pairs amplify sequences that segregate with known disease resistance loci. This method has the advantage of bypassing the subcloning and hybridization steps, thus saving a considerable amount of time. In fact, we were able to map a locus in two days, using this method. We identified four new loci, three of which were linked to loci we had identified by hybridization. None of them however mapped to the blackleg and white rust resistance loci. This suggests that these two resistance genes do not share sufficient homology with *RPS2* to be amplified by our primer sets. A series of primers corresponding to other conserved domains in known R genes should thus be designed and used in different combinations, and in different PCR conditions.

The power of this PCR method can be greatly increased by coupling it with bulked-segregant analysis (Michelmore *et al.*, 1991). Two bulks of DNA from homozygous resistant and homozygous susceptible lines are amplified using a series of primer pair combinations. The PCR products are then digested with various restriction enzymes, and separated by gel electrophoresis. Polymorphisms between the two bulks are likely to be linked to the resistance gene of interest. The disadvantage is that this does not provide a physical anchor to the resistance locus. It is possible however to subclone the polymorphic band, and use it as a probe on a cDNA library. Candidate cDNAs are then

mapped as RFLPs, and those that co-segregate with the resistance gene can be used in complementation experiments to distinguish between the actual resistance gene and members of the same gene family.

2.5 Literature Cited

Adolphe, D. (1980) Canola: Canada's rapeseed crop. Canola Council of Canada. Publication No. 56.

Alarco, A.-M., Balan, I., Talibi, D., Mainville, N., and Raymond, M. (1997) AP1-mediated multidrug resistance in *Saccharomyces cerevisiae* requires *FLR1* encoding a transporter of the major facilitator superfamily. *Journal of Biological Chemistry* **272**:19304-19313.

Allikmets, R., Singh, N., Sun, H., Shroyer, N.F., Hutchinson, A., Chidambaram, A., Gerrard, B., Baird, L., Stauffer, D., Peiffer, A., Rattner, A., Smallwood, P., Li, Y., Anderson, K.L., Lewis, R.A., Nathans, J., Leppert, M., Dean, M., and Lupski, J.R. (1997) A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nature Genetics* **15**:236-246.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**:3389-3402.

Anderson, P.A., Lawrence, G.J., Morrish, B.C., Ayliffe, M.A., Finnegan, E.J., and Ellis, J.G. (1997) Inactivation of the flax rust resistance gene *M* associated with loss of a repeated unit within the leucine-rich repeat coding region. *The Plant Cell* **9**:641-651.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (Eds) (1989) Current protocols in molecular biology. Appendix 3. Greene Publishing Associates and Wiley-Interscience. New York.

Baker, B., Zambryski, P., Staskawicz, B., and Dinesh-Kumar, S.P. (1997) Signaling in plant-microbe interactions. *Science* **276**:726-733.

Bent, A.F., Kunkel, B.N., Dahlbeck, D., Brown, K.L., Schmidt, R., Giraudat, J., Leung, J., and Staskawicz, B.J. (1994) *RPS2* of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. *Science* **265**:1856-1860.

Bisbal, C., Martinand, C., Silhol, M., Lebleu, B., and Salehzada, T. (1995) Cloning and characterization of a RNAse L inhibitor: a new component of the interferon-regulated 2-5A pathway. *Journal of Biological Chemistry* **270**:13308-13317.

Chen, W., Chao, G., and Singh, K.B. (1996) The promoter of a H_2O_2-inducible, *Arabidopsis* glutathione S-transferase gene contains closely linked OBF- and OBP1-binding sites. *The Plant Journal* **10**:955-966.

Dion, Y., Gugel, R.K., Rakow, G.F.W., Séguin-Swartz, G., and Landry, B.S. (1995) RFLP mapping of resistance to the blackleg disease [causal agent, *Leptosphaeria maculans* (Desm.) Ces. et de Not.] in canola (*Brassica napus* L.). *Theoretical and Applied Genetics* **91**:1190-1194.

Doyle, J.F., and Doyle, J.L. (1987) Isolation of plant DNA from fresh tissue. *Focus* **12**:13-15.

Edwards, K., Johnstone, C., and Thompson C. (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research* **19**:1349.

Ferreira, M.E., Williams, P.H., and Osborn, T.C. (1994) RFLP mapping of *Brassica napus* using doubled haploid lines. *Theoretical and Applied Genetics* **89**:615-621.

Ferreira, M.E., Rimmer, S.R., Williams, P.H., and Osborn, T.C. (1995a) Mapping loci controlling *Brassica napus* resistance to *Leptosphaeria maculans* under different screening conditions. *Phytopathology* **85**:213-217.

Ferreira, M.E., Williams, P.H., and Osborn, T.C. (1995b) Mapping a locus controlling resistance to *Albugo candida* in *Brassica napus* using molecular markers. *Phytopathology* **85**:218-220.

Kanazin, V., Marek, L.M., and Shoemaker, R.C. (1996) Resistance gene analogs are conserved and clustered in soybean. *Proceedings of the National Academy of Sciences USA* **93**:11746-11750.

Kemner, J.M., Liang, X., and Nester, E.W. (1997) The *Agrobacterium tumefaciens* virulence gene *chvE* is part of a putative ABC-type sugar transport operon.

Kharbanda, P.D., Ostashewski, M.J., Werezuk, S.P., and McDaniel, R.S. (1995) Update on highly virulent blackleg of canola in Alberta-1995. World Wide Web. http://www.gov.ab.ca/~env/cms/rssd/pestmang/canola95.html

Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E., and Newburg, L. (1987) MAPMAKER: An interactive computer package for

constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**:174:181.

Leister, D., Ballvora, A., Salamini, F., and Gebhardt, C. (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. *Nature Genetics* **14**:421-428.

Lu, Y.P., Li, Z.S., and Rea, P.A. (1997) *AtMRP1* gene of *Arabidopsis* encodes a glutathione S-conjugate pump: isolation and functional definition of a plant ATP-binding cassette transporter gene. *Proceedings of the National Academy of Sciences USA* **94**:8243-8248.

Marrs, K.A., and Walbot, V. (1997) Expression and RNA splicing of the maize glutathione S-transferase *Bronze2* gene is regulated by cadmium and other stresses. *Plant Physiology* **113**:93-102.

Matthysse, A.G., Yarnall, H.A., and Young, N. (1996) Requirements for genes with homology to ABC transport systems for attachment and virulence of *Agrobacterium tumefaciens*. *Journal of Bacteriology* **178**:5302-5308.

Mayerhofer, R., Bansal, V.K., Thiagarajah, M.R., Stringam, G.R., and Good, A.G. (1997) Molecular mapping of resistance to *Leptosphaeria maculans* in Australian cultivars of *Brassica napus*. *Genome* **40**:294-301.

Mgonja, M.A., Dahleen, L.S., and Franckowiak, J.D. (1994) Subsets from mapping populations for localization of new genes in barley. *Barley Genetics Newsletter* **24**:14-23.

Michelmore, R.W., Paran, I., and Kesseli, R.V. (1991) Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences USA* **88**:9828-9832.

Michelmore, R. (1996) Flood warning - resistance genes unleashed. *Nature Genetics* **14**:376-378.

Mindrinos, M., Katagiri, F., Yu, G., and Ausubel, F.M. (1994) The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell* **78**:1089-1099.

Morrow, C.S., Cowan, K.H., and Goldsmith, M.E. (1989) Structure of the human genomic glutathione S-transferase-pi gene. *Gene* **75**:3-11.

Raeder, U., and Broda, P. (1986) Meiotic segregation analysis of restriction site polymorphisms allows rapid genetic mapping. *The EMBO Journal* **5**:1125-1127.

Rubio, N. (1997) Interferon-gamma induces the expression of immediate early genes *c-fos* and *c-jun* in astrocytes. *Immunology* **91**:560-564.

Salmeron, J.M., Oldroyd, G.E.D., Rommens, C.M.T., Scofield, S.R., Kim, H., Lavelle, D.T., Dahlbeck, D., and Staskawicz, B.J. (1996) Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. *Cell* **86**:123-133.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular cloning: a laboratory manual. 2nd Edition. Cold Spring Harbor Laboratory Press.

Sharpe, A.G., Parkin, I.A.P., Keith, D.J., and Lydiate, D.J. (1995) Frequent nonreciprocal translocations in the amphidiploid genome of oilseed rape (*Brassica napus*). *Genome* **38**:1112-1121.

Tenhaken, R., Levine, A., Brisson, L.F., Dixon, R.A., and Lamb, C. (1995) Function of the oxidative burst in hypersensitive disease resistance. *Proceedings of the National Academy of Sciences USA* **92**:4158-4163.

Tomassini, R., Evers, R., Vogt, E., Mornet, C., Zaman, G.J., Schinkel, A.H., Borst, P., and Martinoia, E. (1996) The human multidrug resistance-associated protein functionally complements the yeast cadmium resistance factor 1. *Proceedings of the National Academy of Sciences USA* **93**:6743-6748.

Tomassini, R., Vogt, E., Schmidt, J., Fromentau, M., Amrhein, N., and Martinoia, E. (1997) Differential expression of genes coding for ABC transporters after treatment of *Arabidopsis thaliana* with xenobiotics. *FEBS Letters* **411**:206-210.

U, N. (1935) Genome analysis in Brassica with special reference to the experimental formation of B. napus and peculiar mode of fertilization. Japanese Journal of Botany **7**:389-452.

Wang, G., Warren, R., Innes, R, Osborne, B., Baker, B., and Ronald, P.C. (1996) Construction of an *Arabidopsis* BAC library and isolation of clones hybridizing with disease-resistance, gene-like sequences. *Plant Molecular Biology Reporter* **14**:107-114.

Yu, Y.G., Buss, G.R., and Saghai-Maroof, M.A. (1996) Isolation of a superfamily of candidate disease resistance genes in soybean based on a conserved nucleotide-binding site. *Proceedings of the National Academy of Sciences USA* **93**:11751-11756.

3.0 General Discussion

Natural populations of plants and animals have co-evolved with the microbial world for millions of years. Over time, many microorganisms have developed specialized associations with higher organisms, ranging from symbiosis to pathogenesis. In all cases, elaborate cellular communication mechanisms have evolved in both the microorganisms and their hosts. An increasing amount of evidence suggests that these mechanisms by which a cell senses and responds to its environment are conserved across kingdoms. A striking example is that of disease resistance, which appears to have evolved prior to the divergence of the plant and animal kingdoms, more than a billion years ago (Belvin and Anderson, 1996).

Many parallels can be drawn between the signal transduction cascades via which plant and animal cells recognize and respond to invading pathogens. Likewise, virulence determinants are strikingly similar between plant and animal pathogens. In this context, it is not too surprising to realize that plant disease resistance mechanisms, even to very different pathogens, are conserved throughout the plant kingdom. This homology can be used to identify new resistance determinants, in plants and other species, thereby increasing our knowledge of the interactions between living cells.

3.1 Isolation of Disease Resistance Gene Homologues in Plants

When this project was initiated in 1995, nine genes conferring resistance to viruses, bacteria and fungi had been cloned in mono- and dicotyledonous plant species. Since then, another five resistance genes have been isolated. These fourteen genes belong to five distinct classes, based on homologies in their predicted amino acid sequences

(table 1.1) (Baker *et al.*, 1997). The largest class of plant resistance genes is characterized by the presence of a very conserved nucleotide-binding site (NBS). This homology has been used successfully to identify similar sequences in soybeans and potatoes (Kanazin *et al.*, 1996; Yu *et al.*, 1996; Leister *et al.*, 1996).

The only domain that is present in all but two resistance genes however, is characterized by a series of leucine-rich repeats (LRRs). The only exceptions are *Pto* and *Hm1*. The maize *Hm1* gene does not function in gene-for-gene resistance though, and *Pto* requires a LRR protein, Prf, to confer disease resistance in tomatoes (Salmeron *et al.*, 1996). LRR domains therefore appear to be better indicators of disease resistance genes in general. Because of its very loose consensus sequence however, the LRR domain has not yet been used in the search of resistance gene homologues using PCR. We have attempted to do this by designing primers corresponding to a portion of the *Arabidopsis RPS2* NBS domain, as well as to the more conserved N-terminal portion of the *RPS2* LRR domain. I have shown that these primers produce DNA amplification products that are conserved in the Brassicaceae plant family (see section 2.3.1). While some of these products share little or no homology to genes of known function, a portion of them are very similar to known resistance genes, as well as to different genes involved in other stages of plant and animal disease resistance. These sequences map both singly and in clusters throughout the *B. napus* genome (figure 2.21), as is the case for disease resistance loci in other plant species examined (Kanazin *et al.*, 1996; Leister *et al.*, 1996).

Many of the resistance gene homologues identified in this way are likely to be pseudo-genes (Michelmore, 1996). This is because pseudo-genes accumulate random mutations that are no longer selected for or against. They are therefore more likely to be

polymorphic in a given population. As genetic mapping is based on the analysis of polymorphisms in a population, pseudo-genes will be mapped more readily than functional resistance genes. Indeed, the clone that we have identified that has the most extensive homology to *RPS2* was monomorphic in our mapping population. Likewise, another clone that was found to be homologous to known resistance genes identified 5 different loci on the *B. napus* genetic map and was found to have many stop codons in its sequence. Nonetheless, cloning and mapping these resistance gene homologues is not futile as they provide a physical anchor to resistance loci, and can serve as the ideal markers for the immediate marker-assisted selection of resistance loci in breeding programs (Michelmore, 1996).

3.2 Use of Minimal Populations in Genetic Mapping

We have shown that genetic mapping within chosen intervals can theoretically be achieved using only a subset of the full mapping population (see section 2.3.3). Our results also show the limitations of this method however. For a population subset to be most useful, the population from which it came from must fit two criteria. First, it must display a minimum of segregation distortion. This improves potential genome coverage and the overall resolution power of mapping studies. It is especially significant in this case, as the relative importance of each scoring in a minimal population is greatly increased. Secondly, a saturated genetic map should exist for that population. This allows for the selection of evenly spaced markers that define the mapping intervals of interest. A high confidence level is required in the selection of these intervals, as their importance is also inflated in minimal populations. If these conditions are met, then a rough map

position can quickly and accurately be assigned to any new marker. This method is well suited to the mapping of R-gene homologues, especially if the objective is only to quickly identify areas of the genome putatively involved in disease resistance.

3.3 Quick Mapping of Putative Disease Resistance Genes

We have developed another method to further simplify the quick identification of genomic regions involved in plant disease resistance. This method is particularly useful if the investigator is interested in rapidly identifying sequences that segregate with a known disease resistance locus, rather than in the cloning of a series of R gene homologues. Alternatively, this method could also be used to develop a series of PCR-based markers for disease resistance. These markers could then be used in large-scale breeding programs to aid in the selection of plants with different disease resistance specificities. Our results show that such PCR markers can identify loci closely linked to sequences with known homology to disease resistance genes (see figure 2.21).

3.4 Future Prospects

As more disease resistance genes are cloned and new classes are defined, an increasing number of sequences with homology to these genes will be identified. This will not even require the repetition of studies such as this one, as it is now sufficient to look for similarities in the growing expressed sequence tag (EST) databases to find hundreds of homologous sequences (S. Somerville, personal communication). It will therefore not be long before a very large number of potential resistance genes become available, especially in model plants such as *Arabidopsis*.

To exploit this wealth of information however, it will soon be necessary to develop methods to rapidly test the function of individual sequences (Gibson and Somerville, 1993). In the case of R gene homologues, one possible way is to map all the sequences and determine whether they co-segregate with known resistance loci. This will be of limited immediate use in species, such as *B. napus*, for which little genetic data has been gathered on disease resistance. Another possibility is to identify the corresponding cDNAs and use the latter in two-hybrid screens, using a series of cloned avirulence genes as baits. Such tests would be very powerful, as they would immediately provide information on the resistance specificities encoded by candidate cDNAs.

In any case, it is becoming evident that the cloning of specific disease resistance genes will not be the only use of R gene homologues in the near future. These sequences are likely to also be used in breeding programs for the introgression of chromosomal regions involved in resistance into modern crop cultivars. Such regions have often been lost during the crop domestication process, and breeders frequently need to go back to wild relatives to identify sources of resistance. The availability of a large number of mapped R gene homologues should allow for the introgression and perhaps even the pyramiding, of different sources of disease resistance. In this case it may be advantageous to simultaneously introgress several members of a resistance gene family, even if the function of these other members is unknown, or if they are pseudo-genes. In this way, the natural evolution of the gene family is allowed to proceed, hence allowing the generation of new resistance specificities. This may in fact be more sustainable than the insertion of transgenes coding for specific resistance genes, which may lose their utility once the pathogen has evolved a new virulence mechanism.

3.5 Concluding Remarks

Some aspects of disease resistance appear to have evolved over a billion years ago. The delicate balance that has developed between pathogens and their hosts has ensured the long-term survival of both species. The artificial uniformity created in agricultural systems has upset this balance and results in important crop losses to disease. This is due to the limited adaptability of modern crops with narrowed genetic bases, and to the high selection pressures exerted on the pathogens in such systems.

Increasing the genetic base of crops with various sources of disease resistance simultaneously, and using plant mixtures with different resistance specificities, may help to restore the balance between plants and pathogens. The introgression of several loci at once is very time and resource consuming however. The availability of many mapped resistance gene homologues and their use in the marker-assisted selection of individuals with different combinations of resistance loci should greatly facilitate this task.

3.6 Literature Cited

Baker, B., Zambryski, P., Staskawicz, B., and Dinesh-Kumar, S.P. (1997) Signaling in plant-microbe interactions. *Science* **276**:726-733.

Belvin, M.P., and Anderson, K.V. (1996) A conserved signaling pathway: the *Drosophila* Toll-Dorsal pathway. *Annual Reviews of Cell and Developmental Biology* **12**:393-416.

Gibson, S., and Somerville, C. (1993) Isolating plant genes. *Trends in Biotechnology* **11**:306-313.

Kanazin, V., Marek, L.M., and Shoemaker, R.C. (1996) Resistance gene analogs are conserved and clustered in soybean. *Proceedings of the National Academy of Sciences USA* **93**:11746-11750.

Leister, D., Ballvora, A., Salamini, F., and Gebhardt, C. (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. *Nature Genetics* **14**:421-428.

Michelmore, R. (1996) Flood warning - resistance genes unleashed. *Nature Genetics* **14**:376-378.

Salmeron, J.M., Oldroyd, G.E.D., Rommens, C.M.T., Scofield, S.R., Kim, H., Lavelle, D.T., Dahlbeck, D., and Staskawicz, B.J. (1996) Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. *Cell* **86**:123-133.

Yu, Y.G., Buss, G.R., and Saghai-Maroof, M.A. (1996) Isolation of a superfamily of candidate disease resistance genes in soybean based on a conserved nucleotide-binding site. *Proceedings of the National Academy of Sciences USA* **93**:11751-11756.

Appendix 1: Buffers and Solutions[1]

CTAB buffer

3% (w/v)	CTAB (Sigma)
1.4 M	NaCl
0.4% (v/v)	2-Mercaptoethanol
20 mM	EDTA
100 mM	Tris-HCl pH 8.0

Hoechst Dye Stock

1 µg/µL	Hoechst 33258 dye

Kirby mix

4 g	Tri-isopropylnaphtalenesulfonic acid sodium salt (Phase Separations Inc.)
29 g	4-Aminosalicylic acid (Sigma)
20 mL	1 M Tris-HCl pH 8.0
24 mL	Buffer saturated phenol
360 mL	Deionized H_2O

PCR DNA extraction buffer

200 mM	Tris-HCl pH 7.5
250 mM	NaCl
25 mM	EDTA
0.5%	SDS

TE buffer

10 mM	Tris-HCl pH 8.0
1 mM	EDTA pH 8.0

20X SSC buffer pH 7.0

3 M	NaCl
0.3 M	Tri-sodium citrate

[1] All buffers and solutions are in H_2O, unless indicated otherwise